Integrating Device-to-Device Communications in 5G Cellular Networks

Vlachos, Christoforos

Awarding institution:
King's College London

Integrating Device-to-Device Communications in 5G Cellular Networks

Christoforos Vlachos

A Thesis Submitted for the Degree of

Doctor of Philosophy at

King's College London

July 2018

Dedicated to my parents and my brother.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Vasilis Friderikos for his relentless and insightful guidance throughout my 4-year PhD journey. His passion for technology has motivated me, both consciously and unconsciously, to question everything, seek for innovation and, through hard work, build up my research career.

Special thanks to my second supervisor, Prof. Hamid Aghvami, for his beneficial research directions and advices as well as the head of the Centre for Telecommunications Research (CTR), Prof. Mischa Dohler, who inculcated me with the critical thinking of how research can be translated to entrepreneurship.

I would also like to thank all of my colleagues at CTR, but especially my very good friends Giorgos, Hisham, Kostas, Omar, Ehsan, Gao, Fahad, Yaqub and Massimo for their continuous support and for making my days at King's College London an unforgettable memory.

I gratefully acknowledge the funding sources that made my PhD work possible. I was funded by a EU Multi-Partner Initial Training Network (MITN) Marie Curie project, namely CROSSFIRE, for the first 3 years of my studies. Also, I take the chance to appreciate all of the CROSSFIRE project members for the enjoyable meetings and fruitful discussions we had.

Last but not least, I would like to deeply thank my family for their unconditional love and encouragement. For my parents who equipped me with ethos and kindness as well as supported me in all my pursuits. For my older brother,

Giannis, who taught me how to chase my dreams and be competitive. I would succeed nothing without their continuous support.

Abstract

The evolution of wireless networks towards 5G dictates the integration of a multitude of heterogeneous radio access technologies to the traditional macro-cell systems. Equipping the network with numerous small cell nodes, such as femtocell and picocell base stations (BSs), implies a spectrum efficient and network performance improving solution to support the rapidly increasing user demands. However, this can be proven to be a cost-inefficient method that increases the capital and operational expenditures of the network operators as well as the power consumption, especially in low-traffic network conditions where a number of BSs should be switched-off. To this end, device-centric solutions that leverage the potentials stemming from the proximity, mobility and increased dynamics of user devices should be considered. To this end, direct, proximity-based Device-to-Device (D2D) communication, where two close-ranged user equipments (UEs) are able to exchange data by bypassing the BS, is expected to play predominant role in improving the overall network welfare and ease part of the traffic developed on the BSs side.

This thesis focuses on the soft integration of inband D2D communications in emerging cellular networks where D2D-enabled devices utilize the licensed spectrum. In the introductory part of the thesis we highlight the merits that this communication paradigm can offer in terms of spectrum utilization, energy saving, delay reduction and data rate improvement. We also provide an overview of the D2D use cases that enable opportunities for new services, its potential

in improving the overall network performance as well as its offloading capability that can ease the traffic employed along the network.

In the sequel, we proceed with our proposed methodology that aims at easing the coexistence of cellular and D2D users in emerging cellular networks. One of the main contributions of the thesis is the optimization of cell association for D2D UEs (DUEs). Cell association for D2D communications is an unexplored area and a rather fertile ground for research. Following the conventional motif, a user device would preferably couple with a high power macro cell BS that provides the user with the highest signal power. However, with the advent of D2D communications, this could be proven to be highly inefficient for users that want to communicate directly and are associated with different BSs because BS intercommunication complexity and access delay is introduced. To this end, we propose a number of optimization formulations for D2D-based cell association that takes into consideration not only the nature of the inband D2D communications (underlay or overlay), but also performance-hindering factors such as user density, interference and so on. Other than the throughput enhancing and power saving attributes of the proposed framework, notable resource efficiency improvement is achieved. Indicatively, for both underlay and overlay D2D communications, more than 12% and 45% radio resource utilization mitigation is ensured compared to baseline methods.

On top of optimizing cell association for D2D communications, we further investigate the problem of resource allocation in different D2D underlaying cellular network scenarios where DUEs are permitted to reuse the cellular resources and, therefore, high levels of interference need to be prevented. By considering different deployment scenarios, we propose a set of low-complexity heuristic algorithms with the aim to achieve high data rate performance for D2D communications with respect to meeting the cellular users' quality of service (QoS) requirements. The proposed algorithms are evaluated in high-traffic networking scenarios where D2D communications underlay relay-enabled cellular networks. In aggregate, more than 10% of sum throughput performance is achieved against various resource allocation techniques.

In the sequel, we explore the dynamics of virtualizing the radio resources for efficient sharing as, nowadays, we are witnessing higher network heterogeneity and the emergence of multiple stakeholders with the overarching need to significantly reduce deployment costs and achieve a sustainable network operation. Network virtualization has emerged as a promising technique to overcome the complexity of current network operation as well as facilitate inter-operators' sharing. Therefore, disruptive approaches to manage radio and network-virtualized resources are expected to be a catalyst element of future mobile network architectures. Despite the fact that a number of solutions for radio access network (RAN) virtualization emerged over the last few years, it is worth pointing out that little attention has been placed on issues related to D2D virtualization. Therefore, based on the integration of an inter-tenant controller that enables the radio resource sharing between multiple operators, we devise a set of efficient algorithms to optimize the throughput performance of D2D communications in virtualized environments as well as reduce the utilization levels of the allocated radio resources. More than 12% of sum-rate performance improvement compared to legacy, intra-tenant approaches where the radio resources are assigned based on which device initiates the communication per case.

Finally, a summary of the research outcomes along with some future directions for D2D communications concludes this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

3GPP .... Third Generation Partnership Project

4G .... Fourth Generation

5G .... Fifth Generation

API .... Application Programming Interface

BBU .... Baseband Unit

BN .... Buffon's Needle

BS .... Base Station

BW .... Bandwidth

CAPEX ...... Capital Expenditures

CAS .... Cell Association

CbH .... Cost-based Heuristic

CDF .... Cumulative Distribution Function

CR .... Crossing Ratio

C-RAN .... Cloud - Radio Access Network

CSD .... Caching-Server Device

CSI .... Channel State Information

CUE .... Cellular User Equipment

D2D .... Device-to-Device

dB .... Decibel

DD .... Disjoint Decoupled

DL .... Downlink

DUE .... D2D User Equipment

EU .... European Union

FD .... Full Duplex

FIFO .... First-In First-Out

FFR .... Fractional Frequency Reuse

FRF .... Frequency Reuse Factor

FTP .... File Transfer Protocol

GA .... Genetic Algorithm

HD .... Hybrid Decoupled

HetNets ...... Heterogeneous Networks

HRS .... Heuristic Resource Slicing

ICIC .... Inter-Cell Interference Cancellation

ILP .... Integer Linear Programming

InP .... Infrastructure Provider

i-RRA .... iterative Randomized Resource Allocation

ISM .... Industrial, Scientific and Medical

ITU .... International Telecommunication Union

JC .... Joint Coupled

JD .... Joint Decoupled

L3 .... Layer-3

LRU .... Least Recently Used

LS-CRRM ... Large-Scale Cooperative Radio Resource Management

LTE .... Long Term Evolution

LTE-A .... Long Term Evolution - Advanced

MCS .... Modulation Coding Scheme

MINLP ...... Mixed Integer Non Linear Programming

MNO .... Mobile Network Operator

MOCA .... Multi-Objective Cell Association

MS .... Mobile Station

MVNO .... Mobile Virtual Network Operator

NC .... Network Coding

NFV .... Network Function Virtualization

NSI .... Network State Information

NVS .... Network Virtualization Substrate

ONF .... Open Networking Foundation

OP .... One-point

OPEX .... Operational Expenditures

ORS .... Optimal Resource Slicing

OTN .... Optical Transport Network

P2P .... Peer-to-Peer

PPP .... Poisson Point Process

PRB .... Physical Resource Block

QoS .... Quality of Service

RA .... Resource Allocation

RAN .... Radio Access Network

RB .... Resource Block

RRH .... Radio Remote Head

RSS .... Received Signal Strength

SC-FDMA ... Single Carrier Frequency Division Multiple Access

SDN .... Software-Defined Networking

SINR .... Signal-to-Interference-plus-Noise Ratio

SIR .... Signal-to-Interference Ratio

TP .... Two-point

UE .... User Equipment

UL .... Uplink

V2V .... Vehicle-to-Vehicle

VFM .... Virtual Function Manager

VIM .... Virtualization Infrastructure Manager

vNF .... virtual Network Function

VoIP .... Voice over Internet Protocol

WLAN .... Wireless Local Area Network

WRV .... Wireless Resource Virtualization

List of Symbols

α .... Path loss compensation factor

$\gamma_{()}^{()}$ .... SINR ratio

Γ .... Upper limit of usable resources

$\mathcal{B}$ .... Set of base stations

$\mathcal{C}$ .... Set of cellular users

$\mathcal{K}$ .... Set of resource blocks

$\mathcal{L}$ .... Set of D2D links

$\mathcal{M}$ .... Set of relays

B_{RB} .... Bandwidth of resource block

$G_{()}^{()}$ .... Gain factor

$I_{()}^{()}$ .... Interference factor

$P_{()}^{()}$ .... Transmit power

$PL_{()}^{()}$ .... Path loss

$R_{()}^{()}$ .... Achievable rate

$s_{()}^{()}, \tau_{()}^{()}, x_{()}^{()}, y_{()}^{()}, z_{()}^{()}$ Indicative decision variables for the optimization problems

* The above List of Symbols is restricted to the basic notations that are followed throughout the Thesis. Also, the blank parentheses mean that the above symbols can be enhanced with different per case (e.g. problem formulation) symbols in subscript and superscript

Chapter 1

Introduction

The unquestionable need to address the rapidly increasing and varying demands for wireless access to the existing networking infrastructures is the main concern for network operators. The resulting mobile data traffic, most of which stems from video requests, is expected to further challenge the network capacity limits and grow almost eightfold by the end of 2020 [1]. This growth highly correlates to the ongoing user proliferation as well as the evolution of mobile devices. Devices with increased set of capabilities, such as smartphones and tablets, create a fertile ground for introducing new business models and propelling a multitude of novel services, including content sharing, social networking services [2], advertising and multiplayer gaming, where, among all, users are expected to interact more in a local, proximity-based fashion.

On top of the increasing capacity that network operators have to cope with, other significant performance requirements such as latency, throughput, reliability and energy efficiency need to be ensured. This is the reason why cellular networks have been gradually experiencing radical architectural changes through the years [3]. Moving from the traditional, macro-cell only networks to more flexible architectures that network heterogeneity introduces, several radio access technologies are being ongoingly integrated in order to improve the overall network sustainability [4]. This shift stably prepares the ground for migrating towards the next generation mobile networks, referred to as the fifth generation (5G). 5G networks are envisaged to rely on multi-tier architectural designs where the user devices will be playing a predominant role by operating on different modes and connecting flexibly to different nodes [5][6].

However, the continuous embodiment of different access technologies such as small cells and other low-power nodes as well as the installation of macro base stations (BSs) to increase capacity and support a huge number of connections can become cost-inefficient and eventually lead to the increase of capital and operational expenditures (CAPEX/OPEX) for the network operators' side. Hence, device-centric solutions that exploit the merits of existing technologies have to be further enhanced. To this end, Device-to-Device (D2D) communication has arisen as a disruptive technological paradigm and an additive device tier to existing macro-cell based systems, especially in order to meet the locality-based service demands as well as benefit the overall cellular network's operability [7]. Briefly, in this communication paradigm, two close-ranged user equipments (UEs) are able to exchange data between each other by establishing a direct link and bypassing the routing of data through the BS [8]. The principal driving force for adopting such a technology is the inherent potential for mobile stakeholders to offload part of the traffic developed to the core network as well as the improvement in terms of network resource efficiency, power savings and the received quality of service (QoS) of the subscribed users, especially for proximity-based connections.

In this chapter, the significance of enabling the D2D communication paradigm in cellular networks is highlighted. Initially, in section 1.1 the focus is turned on indicating the technical as well as commercial suitability of D2D in current Long Term Evolution (LTE) based networks and beyond. Then, section 1.2 discusses the necessity of achieving efficient D2D-aware cell association for the paired communicating devices and the overall impact on conventional cellular connections.

Furthermore, resource allocation for D2D underlaying cellular networks is a critical aspect that needs to be addressed in order to ameliorate the achieved data rates with respect to interference patterns within the network and is a matter of interest in the same section. Finally, the main contributions and outline of the thesis are described in 1.3.

1.1 Device-to-Device Communications

One of the principal commitments of the Third Generation Partnership Project (3GPP) LTE networks was the unprecedented increase of achievable data rates and system capacity [9]. Then, its descendant, LTE-Advanced (LTE-A), commercially known as 4G, was further envisaged to support new technologies and components for LTE in order to meet even more challenging communication demands in a cost-efficient way. Higher bitrates, enhanced spectral efficiency as well as capacity to accommodate an increased number of simultaneously active subscribers were set by the international telecommunication union (ITU) as the most significant requirements [10].

To this direction, an important aspect that is able to offer multiple benefits and needs to be exploited is that of local area communications, specified for realizing short-range connections between the transmitter and the receiver, which are able to achieve disruptive data rates with low energy consumption. For this reason, existing technologies including wireless local-area network (WLAN) systems based on the standards of IEEE 802.11 (e.g. Wi-Fi, Wi-Fi Direct) that are able to provide local services and fast access to the Internet through the license exempt bands (e.g. industrial, scientific and medical (ISM) bands) have been broadly used and consequently attracted lot of research interest [11]. However, the use of unlicensed spectrum renders the controllability of interference a hard task for the local service providers and creates inconvenience whether they can guarantee a stable, controlled networking environment or not. Therefore, realizing wireless communications in the licensed bands of a cellular network can entail a more controllable, interference-limited environment where all communication types will be under the control of the BS or other central entities.

This is the reason why D2D has become a popular item of discussion and industrial other than academic interest. Initially, D2D related communication setup principles were defined within 3GPP Rel. 12. In specific, 3GPP Rel. 12 and Rel. 13 elaborated on the proximity services (ProSe) enablement, based on the D2D operation, and also highlighted the potentials and evolution capabilities of the concept of D2D to Vehicle to Vehicle (V2V) communications. 3GPP Rel. 13 also referred to multi-hop relay networking capabilities of D2D communications as well as presented the standardization efforts with regard to exerting priority control for mission critical push-to-talk applications. Then, the discussion under 3GPP Rel. 14 constituted a precursor of the 5G network development endeavours and focused on a multitude of aspects that included among all emergency services, LTE-and-beyond enabled services and functionalities for vehicle-to-everything (V2X), location services and latency reduction techniques.

With the advent of 5G, which is expected to bring phenomenal changes to the current state of cellular networks, D2D communication is envisaged to become the key enabler of proximity-based, direct communications operating in the licensed band and can be thought of as the cellular-based peer-to-peer (P2P) equivalent of Wi-Fi Direct. Different concepts of enabling D2D into future networks exist, and can be mainly classified into two categories [11]:

- *Inband D2D*: In this category, D2D users are eligible to utilize the same licensed spectrum that is available for cellular UEs (CUEs). The main advantage of it is the high control over the cellular spectrum that network operators can achieve. This translates to potential interference avoidance which is not the case in the hard-to-manage unlicensed spectrum. The inband category is further divided into two subcategories: first, the *underlay*, where D2D and cellular users use the same spectrum. This increases the spectral efficiency in the network. Second, it is the *overlay* case where D2D communication is allocated with dedicated part of the cellular spectrum. In the latter, even though interference between D2D and cellular users is eliminated, the spectral efficiency of the cellular spectrum decreases. In addition, the challenge of deciding which part of the spectrum needs to be allocated for the direct communications is still an open and hard to solve issue.

- *Outband D2D*: In this category, D2D communication exploits the unlicensed spectrum. The incentive for introducing such a concept is the elimination of interference developed between D2D and cellular links. However, as already mentioned, the unlicensed bands, such as ISM, create inconvenience in terms of interference controllability. Thus, there is no network control over D2D communications. In the unlicensed spectrum, usually other wireless technologies such as Zigbee, Wi-Fi Direct or bluetooth are adopted in order to realize outband direct communications [11]. Outband D2D is also further categorized into *controlled* and *autonomous* communication. In the former, it is suggested that the control of the second interface is given to the cellular network, whereas in the latter D2D is used autonomously by the users as the second interface and is not under the control of the cellular network.

The main difference between D2D integrated in cellular networks and other direct technologies is that the former is subject to a set of QoS rules mandated by the Telecom operator, unlike the rest that are based on "best effort" communications. In this thesis, the focus is turned on the inband nature of D2D communications where the users are supported by the cellular infrastructure and are allocable with resources only from the licensed spectrum. In specific, we mainly consider the most popular D2D communication type, that of the underlay due to its high spectral efficiency attribute as a consequence to the reuse gain it offers. Practically, this implies that the radio resources' reuse factor is tightened in a way that a radio resource block (RB) can be used more than once within the same cell. Additionally, the proximity of two close-ranged UEs leads to extremely high data rates, low latency as well as controllable transmit power and, thus, low energy consumption. If we also consider that D2D communication is an economical technology that uses the pre-existing cellular infrastructure and is able to alleviate and ease part of the traffic developed in the network, network operators are considering introducing it in order to enable opportunities for new locality based applications and improve current services throughout the cellular network [12][13].

1.2 Cell Association and Resource Management

This thesis aims at optimizing two critical networking issues, that of cell association and resource allocation in D2D-enabled cellular networks with regard to a number of inhibitory constraints, such as interference and resource availability limitations. The following subsections highlight the need for devising novel techniques that take into consideration the D2D communication's nature in order to utterly exploit the offered benefits.

1.2.1 The Need for D2D-aware Cell Association

Conventional cellular networks are characterized by the deployment of multiple homogeneous macro BSs with similar transmit power magnitudes and regular traffic loads. User devices in such networks associate with the BS that serves the geographical cell area that they belong to. This technically means that a UE connects to the BS that provides the best signal-to-interference-plus-noise ratio

(SINR) and establishes both downlink (DL) and uplink (UL) connections that carry control and data traffic. If a uniform distribution of the users is observed within a multi-cell network, this implies that each BS will be approximately serving the same number of subscribers and, therefore, the cellular connections are efficiently orchestrated without the need of load balancing.

However, in recent years, the evolution of mobile telephony as well as the deluge of mobile devices with enhanced capabilities evoked the migration from the homogeneous to heterogeneous networks (HetNets) that would play a significant role on facing the immensely increasing capacity demands. The amalgamation of multiple access technologies and the deployment of base stations with different physical sizes as well as transmit powers represent the notion of HetNets that was standardized within the 4G specifications. In specific, the LTE-A standards designed a multi-tier HetNet roll-out, where macro cells would be overlaid by small cells, such as picocells, femtocells, and relays. The aim of injecting the small cell technology was to mainly offload the traffic developed on the macro BSs' side and to improve the coverage quality as well as the overall performance of mainly the cell-edge users. Therefore, this cell densification, where different base station types with varying transmit powers and capabilities co-exist, complicates the issue of cell association and creates ambiguity considering where the user devices should be connected to.

Although the aforementioned installation of small cell BSs, especially in high-traffic, hotspot zones within the network has been proven to be an efficient method that improves capacity as well as coverage, the irregularity of the spatial distribution of user devices is a repressive factor to the proper balancing of the network traffic load. Based on the norm of associating a user with the maximum SINR-providing BS in the downlink, this can potentially create huge imbalance cases while most of the UEs are being coupled with the macro BS that serves the cell area they are located in. This translates to increased underutilization effects for the small cells and a potential bottleneck for the macro cell infrastructure. It also indicates that the 4G architecture was not effectively designed to support the integration of these new technologies because the network performance is burdened with further enhancements, in that case in the form of cell association related additions. To this direction, several methods have been proposed throughout the literature and applied within the industry to allow for either avoiding *a priori* such imbalance cases or coping with it *a posteriori*, and will be presented in the sequel.

On top of that, even though D2D communication is expected to constitute a significant part of future, wireless based connections, little attention has been paid on how the association for two proximate and cooperating users should be taking place. In such case, it is obvious that two communicating UEs will be associated with either the same or different BS according to the rule of maximum SINR. However, in the case of different associated BSs for a pair of two UEs, the latency relating to BS intercommunication for exchanging information about the D2D link establishment might be prohibitive. Hence, both UEs might be preferably connected with the same BS in order to facilitate the synchronization procedure.

Also, the ongoing user densification combined with the irregularity of the cells' shape and the co-existence of multiple user types can lead to different levels of load congestion in the cell areas (from lightly loaded to severely congested cases) which need to be effectively managed in order to increase network capacity [14]. To this end, and by taking into consideration the nature of D2D communication, intelligent cell association and traffic balancing techniques must be applied to address the resource and capacity limitations which might lead to network bottleneck phenomena.

Related work

As already discussed, cell association in cellular networks has been a critical challenge for network operators due to the ongoing increase of user demands in an underlying resource-limited infrastructure. The scope of further improving existing cell association techniques is to enhance network capacity, accommodate more users simultaneously with respect to their QoS requirements as well as allow for the integration of other communication types, such as D2D communication.

Up to date, different approaches for cell association, mainly concerning the traditional cellular connections in homogeneous and heterogeneous networks, exist within the literature. An exemplary work that considers both cases is [15]. It proposes a base station - mobile station (BS-MS) association policy based on the highest signal-to-interference ratio (SIR) where each MS should be located within a predefined maximum distance from the BS to be served by it. The integration of the maximum distance limitation accounts for avoiding frequent handovers. They also study the same problem when distance is set to infinite, and thus highest SIR is the association criterion for single-tier (homogeneous) networks. Also, they further apply the same approach for HetNets and prove its applicability for user offloading to lightly loaded cells. However, no attention has been paid on how the association of communicating D2D users should be realized.

Regarding HetNets, cell association has been mainly based on the downlink received signal strength (RSS) estimations of the cellular users. The integration of different access technologies such as pico/femtocells that operate over the same spectrum as that of the underlaid cellular network introduces another degree of complexity compared to homogeneous settings, due to the developed inter-cell interference. A number of solutions to encounter this problem have been applied, including, inter alia, cell splitting, range expansion, semi-static resource negotiation on third-party backhaul connections, and fast dynamic interference management. Those schemes and their respective benefits are well presented in

[16]. However, because of the transmit power disparities of different tiers (macro-cells and small-cells), imbalanced association cases will be appearing as most of the users will be coupled with the macro BS [17]. In order to encounter this issue, several works leveraged the concept of cell biasing, where the power signal that a UE receives from a deployed small cell is increased by adding a biasing factor [18]. With this method, cell association imbalances can be reduced. Its basic benefit is that network capacity improvement is achieved via its macro-to-small cell offloading attribute [19]. However, this benefit is followed by an associated drawback, especially in highly dense scenarios; this bias-centric user association might lead to unexpected interference patterns as the biased users will receive interference from the nearby macro-cell [20]. This happens due to overlapping radio resource availability between the two adjacent cells where each of the corresponding base stations controls the resource allocation patterns for only the users that are associated with it. Further, a comprehensive SINR analysis aiming at estimating a user's association probability, outage probability as well as the spectral efficiency based on flexible cell association with different BS types (e.g. macro or picocell BSs) is studied in [21]. Therein, a set of numerical results has proven that there might be some cases where the random addition of pico and femtocells to a cellular network will not necessarily increase the network capacity and overall welfare. Extensively, an analytical taxonomy of cell association techniques is detailed in [22].

Due to the aforementioned issues, [23] studied the effect of decoupling downlink (DL) and uplink (UL) sessions in dense HetNets and illustrated the substantial performance gains in the UL for real world scenarios. This decoupling notion differentiates the way UEs get connected with a BS. In downlink, cell association is based on the downlink received signal power, whereas the uplink cell association depends on the path-loss estimations. Not only UL throughput could be significantly improved, but also outage rates are decreased while ensuring a minimum throughput requirement. UL/DL decoupling also provides a cell association-based insight of how future 5G systems could be implemented to ensure improved performance for both sessions [24]. Thus, the UL based CAS should be carefully designed and updated as new technological components such as D2D communications are expected to become a significant proportion of overall connections in cellular networks.

Even though the integration of D2Ds in cellular networks brings up a number of challenges and merits, cell association for this communication paradigm has been barely studied. In [25], the authors proposed an efficient four-step load balancing mechanism when service-requesting users are associated with fully congested cells. The aim there is to transfer part of the developed traffic to the less congested cells in a multi-tier network by making use of relay-enabled direct (D2D) communication. Traffic imbalance phenomena have to be taken into account as, in many cases, the need for direct communication emerges in a rather irregular fashion in space.

Then, in [26], the authors developed a joint framework that considers the user association and transmission mode switching between direct and D2D relay modes in order to improve spectral as well as energy efficiency via closed-form solutions.

1.2.2 Resource Management for D2D Communications

The most popular among all the 3GPP defined D2D communication modes is that of inband. It introduces a network authorized direct communication where both UEs are being served by LTE infrastructure and are able to utilize the cellular resources in order to communicate between each other. This use of the cellular spectrum for D2D other than the conventional cellular connections entails a more interference-controllable and spectrum efficient environment for network operators. However, depending on the spectrum sharing case of either underlay or overlay D2D communication type, network operators have to adopt different approaches on how to orchestrate the scarce radio resources among cellular and D2D links.

On the one hand, the integration of D2D as an underlay in cellular future-based networks inserts another degree of complexity to the resource management issue due to the potential simultaneous reuse of the available radio resources from cellular and D2D links. In such case, high levels of mainly intra-cell interference caused by the same resource reuse between a cellular and a D2D UE may arise. Also, non-negligible inter-cell interference stemming from multiple connections has to be taken into account. To this direction, novel and disruptive resource allocation techniques have to be devised in order to not only optimize the performance of D2D communications but also improve the overall network efficiency by limiting the developed interference.

On the other hand, the notion of overlay D2D communication is based on assigning part of the cellular, licensed spectrum only for D2D links. Thus, dedicated resources are allocated for satisfying the D2D connectivity needs. This creates orthogonal available resource pools for cellular and D2D users which consequently eliminates the interference exerted between each other. Taking into account that cellular users are also being assigned with orthogonal resources, two types of interference may emerge: 1) inter-cell interference only among cellular connections or D2D links, and, 2) intra-cell interference stemming from same resource use by multiple D2D links. It is obvious that the main advantage of overlay D2D is the limited interference compared to the underlay. However, the radio resource orthogonality that is introduced comes at the cost of spectrum efficiency and the spectrum proportion to be allocated for D2D communication is in principle network traffic-dependant and is hard to be decided.

D2D communication, even though able to exploit either DL or UL resources of a cellular system, is more likely it will be operating over the less utilized UL;

this admission is also followed in this thesis. This means that in underlay a D2D transmitter may cause interference to the BS, which is more interference-proof and has stronger processing abilities than the user devices [27]. Most of the destructive interference derives from closely located cellular transmitters and D2D receivers that utilize the same spectrum. On the contrary, in the overlay scenario, different D2D pairs that are using the same radio resources can be exerting interference among each other, while only inter-cell interference for the cellular links exists.

Resource allocation for D2D communication is a well-investigated subject, especially as an underlay in cellular networks. However, existing methods to improve or even optimize the resource sharing aspect between D2D and cellular users have to be further improved and extended in order to be applied to varying network settings. To this end, disruptive resource management techniques that take into consideration not only traditional homogeneous networks, but also high-traffic HetNets or even virtualized environments have to be devised. The main scope is to optimize the performance of D2D communications while cellular connections are being negligibly harmed and resource efficiency is not violated.

Related work

In order to harvest the potential gains that D2D communication can offer in cellular networks, it is of critical importance to properly design peer discovery mechanisms, physical layer procedures and actions as well as resource management techniques to increase the networks' spectrum efficiency [13][28]. Assuming that the two former follow the 3GPP standardization efforts and can be specified as in [11], in this paragraph we focus on reviewing existing resource management mechanisms for D2D communications in cellular networks.

One of the kick-off works on the management of D2D communications is cellular networks is that of [29]. The authors proposed a resource allocation scheme for reducing the intra-cell D2D-to-DL and UL-to-D2D interference to an from cellular links, respectively. It is an indicative work on the merits of introducing local P2P direct connections in comparison to the case of routing the data traffic via the BS. More than 2.3 times median cell capacity increase was yielded with the proposed resource sharing and interference-aware method. In a similar fashion, [30] addresses the problem of spectrum sharing between cellular and underlaying D2D communications and analyzes the interference caused by D2D transmissions to the cellular connections in both UL and DL phases. Because the resource allocation problem in a D2D underlaying cellular network scenario is a mixed integer non linear problem (MINLP), a greedy heuristic algorithm which restricts the exerted interference by utilizing channel gain information from the cellular connections is alternatively devised.

Another interference-aware scheme for mitigating the problematic near-far interference in D2D communications as an underlay in cellular networks is proposed in [31]. The aim is to carry out a near-far interference-aware method of assigning D2D resources based on time hopping as well as optimize the time-hopping parameter settings in order to improve the overall network performance in terms of varying service requirements. In a similar scenario, in [32], the authors propose a distance-constrained resource sharing criterion that facilitates the BS to choose a cellular user's resource for a D2D pair and prevents the interference from surpassing a certain level. The proposed criterion's applicability is demonstrated based on two different power control schemes and evaluated in terms of outage probability for D2D communications.

The unconventional assumption of reusing more than one cellular user's resources to serve a D2D link's service requirements is extensively followed within the D2D related literature. The reason behind this assumption is to not only increase the data rate performance of D2D communications but also improve the spectrum efficiency of the network.

Based on this assumption, different algorithmic and mathematical approaches are followed. In [33] the idea is to prioritize the cellular user and then apply a resource allocation optimization solution for D2D communications that underlay a cellular network. The simulation results showed that the proposed method performs better compared to the case that a D2D link can only reuse one cellular user's resources in both UL and DL. However, in a high-traffic scenario where multiple cellular and D2D connections might exist, this method is not necessarily applicable as resource availability decreases. The same assumption on reusing multiple users' cellular resources for meeting the D2D-related QoS demands (in specific, FTP and VoIP services) is used in [34]. The authors propose a two-step resource allocation scheme. Firstly, each D2D pair, being eligible to choose among several candidate RB groups that consist of available radio resources (the number of which depends on the target QoS and the link quality of the two users on these RBs), is assigned with the RB group that utilizes the least number of resources. At the same time, the QoS of the cellular users needs to be ensured. However, in case of high D2D traffic, this would eventually lead to depletion of the available RB pool or insufficient allocated resources to meet the D2D QoS targets.

Game theory is a very popular mathematical model that is massively used in communications-related problems and is mainly applicable to non-centralized, distributed networking solutions. An exemplary work that introduces a sequential second price auction as a novel DL resource sharing scheme for D2D and cellular communications can be found in both [35] and [36]. The authors firstly formulate the value of each radio resource unit for each D2D link and then introduce a N-ary tree as the basis of their auction-based algorithm. In the aforementioned works, different simulation results showed that this algorithm outperforms baseline resource allocation methods in terms of system sum rate, efficiency and fairness. Further, in [37] a Stackelberg game framework where a cellular user and a D2D link form a leader-follower pair is developed. In such case, the role of the follower is played by a D2D UE who “buys” channel resources from the leader (cellular UE). This constitutes the basis of their proposed algorithm for realizing joint time scheduling and radio resource allocation for D2D communications in cellular networks. Finally, [38] proposes a game-theoretic framework that achieves a tradeoff between energy efficiency and interference in D2D underlaying cellular networks' scenarios. In specific, a distributed coalition formation algorithm is developed based on merge-and-split rule and the Pareto order.

The selection of communication mode (cellular or direct mode) and resource sharing scheme (non-orthogonal or orthogonal) is addressed in [39]. Therein, the authors contemplate the optimization of the aggregate throughput over the shared resources subject to spectrum and power restrictions. To this direction, two different optimization approaches are considered. First, a greedy sum-rate maximization scheme with respect to a maximum energy consumption threshold is devised, where D2D and cellular users are viewed as competing services. Then, a sum-rate maximization problem subject to rate constraints and based on prioritizing the cellular users to guarantee their service demands is proposed. However, the proposed prioritization might constrain the D2D communications' performance by allocating them with the remaining, more interference-prone radio resources. The achievable sum-rate is limited by practical restriction deriving from the considered modulation and coding schemes (MCS). Lastly, the authors provide a thorough representation of the advantages and disadvantages of the compared resource sharing methods in terms of throughput and power performance.

However, most of the works have been focusing on using the UL resources for satisfying the D2D communication needs since in UL less traffic is encountered and the D2D transmitters' power does not strongly affect the cellular connections. In [40], the authors initially introduced an interference coordination strategy and followed up with a channel reuse selection optimization technique in a single cell scenario which aims at maximizing the number of simultaneously served D2D connections and minimizing interference caused by D2Ds. An optimal solution is sought by relaxing the MINLP resource allocation problem and solving it via a Hungarian algorithmic proposal; finally a low-complexity heuristic algorithm is devised in order to achieve a near-optimal performance. In a more complex network setting, the optimal D2D spectrum allocation over multiple bands in heterogeneous networks is analyzed in [41]. The objective is to schedule D2D pairs to different frequency bands and improve the capacity performance for D2Ds. Again, the problem is translated to a solvable version via Lagrangian relaxations, which however entails a computationally complex algorithm.

In a try to optimize the sum-rate performance for D2D communications by guaranteeing the QoS demands of cellular users, the authors in [27] analytically characterize the optimal resource sharing that maximizes the D2D throughput while retaining the cellular performance in acceptable levels. Due to the non-convexity of the problem, they reduce it and propose two alternative sub-optimal solutions by relaxing basic power and radio resource related constraints with regard to cellular users. Even though their solutions provide a near-optimal performance, there are two drawbacks: first, the high complexity of the provided solutions, and second, that it is restricted in a single-cell scenario. Resource sharing optimality is of interest also in [42] where the authors introduce a bipartite matching strategy based on graph theory in order to match a group of D2Ds with a cellular UE (share the same resources) that minimizes interference. This method outperforms the random allocation case and provides a reduction in terms of computational complexity as compared to the optimal case.

Other than the direct single-hop communication between two users, a connection can be established by far-located peers through multiple hops. This premises that the intermediate nodes are relay-eligible to enable routing the data traffic flows from one side to the other. The concept of relaying, previously applied in ad hoc networks with separate frequencies [43], has attracted a lot of interest in cellular networks where user devices can communicate between each other or with the BS through multiple, direct D2D links [44]. The main hindering aspects of its application is the spectrum utilization (how the resources will be allocated in time and frequency) and the interference involving cellular and D2D communications. In [45], a practical discovery protocol that allows for the establishment of a route between intended-to-communicate users in a distributed fashion is proposed. The authors' aim is to minimize the outage probability and to reduce the number of transmission required to discover a path to the destination node. However, in such distributed networks, such D2D-only related improvement might come at the cost of cellular connections and specifically in the form of massive cumulative interference to the BS side. Other similar distributed solutions are proposed in [46] and [47]. Based on the placement of multiple layer-3 (L3)[1] relays in space, the authors propose an optimization technique for allocating radio resources at the relays in order to maximize the end-to-end throughput as well as guarantee that cellular and D2D users will satisfy their QoS requirements. Even though of high importance, the authors have not considered end-to-end delay and latency as critical QoS parameters for achieving far, relay-assisted communications.

1.3 Contributions and Thesis Outline

This thesis mainly aims at optimizing two challenging networking issues, that of cell association and resource allocation in D2D-aware cellular ecosystems. A description of the main objectives and contributions is elnisted below.

- The integration of D2D communication paradigm in future wireless networks adds another degree of complexity considering cell association in cel-

[1]the L3 relay incorporates the same functions as a cellular base station and is able to eliminate noise and inter-cell interference

lular networks. In specific, ambiguity is inserted regarding whether the two communicating D2D nodes should be associated with the same BS or different ones. To this end, an extensive study considering the merits and drawbacks offered by associating the communicating DUEs with the same or different BS is provided in Chapters 2 and 3. In specific, integer linear programming solutions for the cases of D2D overlay and underlay scenarios are proposed and then validated by comparing its performance with baseline existing methodologies. Part of the Chapter 2 has constituted the basis for the publication *"Interference-Aware Decoupled Cell Association in Device-to-Device based 5G Networks"* shown in the Appendix, whereas *"MOCA: Multi-Objective Cell Association for Device-to-Device Communications"* contribution is the basis for Chapter 3.

- Furthermore, we investigate the problem of resource allocation in different D2D underlaying cellular network scenarios where DUEs are allowed to reuse the licensed resources, traditionally allocated for cellular UEs (CUEs). The reason why we explore this subcategory of the D2D taxonomy is because, other than being the most spectrum-efficient D2D communication type, it is also the most challenging as high levels of interference, exerted to/from cellular users due to frequency reuse patterns, need to be controlled. Taking into account that multiple user connections such as cellular, D2D and even relaying might exist, in Chapter 4 we propose a set of optimization formulations as well as alternate low-complexity heuristic algorithms aiming at achieving high data rate performance for D2D users while at the same time the cellular users' performance and QoS is respected. Disruptive performance gains compared to existing resource allocation techniques prove the supremacy of our proposed framework in terms of throughput and spectrum efficiency. The paper *"Bio-Inspired Resource Allocation for Relay-Aided Device-to-Device Communications"* included in the Appendix section has been the basis of Chapter 4.

- In Chapter 5 we study the issue of resource allocation for D2D communications from a more futuristic point of view by exploring the dynamics of virtualizing the scarce radio resources for efficient sharing among different network operators and multiple stakeholders. The enabler of it is the notion of network virtualization which recently emerged as a novel concept to overcome the complexity of current network operation as well as facilitate inter-operators sharing. Following the vision of RAN virtualization to manage radio and network virtualized cellular resources, we introduce the concept of an inter-tenant (inter-operator) controller that enables resource sharing between multiple operators with respect to the existence of D2D communications in future-based cellular networks. The idea behind it is that two or more network operators sharing the same infrastructure are willing to offer part of their allocated radio resources for the satisfaction of D2D communication needs under the control of the aforementioned inter-operator entity. Again, the criterion for evaluating this proposal is the rate performance achieved by D2D and cellular users and is compared with RAN virtualization related works. Chapter 5 contributions have been published under the titles *"Optimal Virtualized Resource Slicing for Device-to-Device Communications"* and *"Optimal Virtualized Inter-Tenant Resource Sharing for Device-to-Device Communications in 5G Networks"* as presented in the Appendix section.

- Finally, a wrap-up of the research findings along with future directions and insights considering the integration of D2D communications in future networks is given in Chapter 6.

Chapter 2

Overlay D2D Cell Association Optimization

2.1 Introduction

The booming cellular network traffic growth over the years has initiated the migration from single-tier homogeneous networks to multi-tier HetNets era in order to face the immense capacity demands in hotspots efficiently and scalably. The HetNet solution not only leads to increased network capacity but also brings the network closer to the users' side. On top of that, D2D communication introduces similar merits that mainly derive from the proximity of user devices and by enabling direct communication between each other without the need for routing the data via the fixed infrastructure network [48].

As already discussed, until the fourth generation of cellular networks cell association (CAS) has been based on the DL received signal power only. However, in [23] the authors proved that a UE's association with a BS in both UL and DL sessions based on the DL received power in a HetNet can be highly suboptimal. The idea of decoupling UL and DL resulted in phenomenal gains in the UL case. On the other hand, D2D UEs are enabled to operate in both cellular and direct

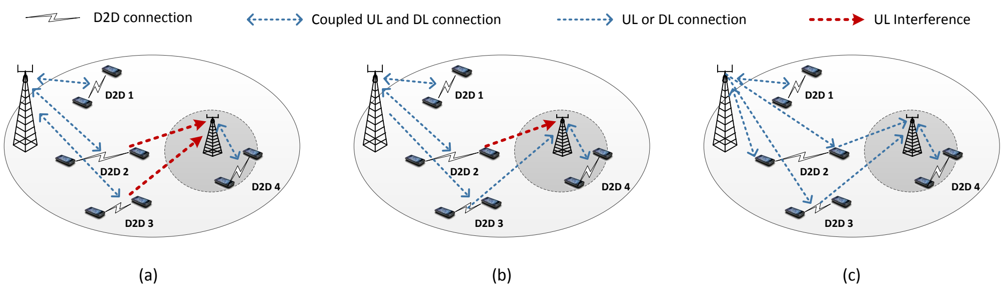


Figure 2.1: Considered cell association scenarios: (a) Joint-Coupled, (b) Joint-Decoupled, (c) Disjoint-Decoupled.

modes in subsequent time instants or subframes. Hence, D2D cell association needs to also consider the nature of cellular transmission. According to the 3GPP specifications [8], D2D communication will be operating in the UL licensed band which makes the idea of decoupled association closely linked to the D2D cell association problem. Even though cell association has been a matter of interest in macro-cellular systems, only recently has been thoroughly studied in HetNet settings [49]. However, as D2D is expected to constitue a big portion of future wireless connections, D2D-related cell association has been barely studied and needs to be well established [50].

The aim herein is to investigate the issue of cell association for inband D2D communications in a heterogeneous network and by conidering different CAS strategies. D2D and decoupled UL and DL have been both envisaged to become principal building blocks in future 5G networks [51]. In this chapter, the focus is initially turned on the inband overlay communication where D2D and cellular communications both take place in the licensed band but are assigned to operate over different frequency subbands. The contribution in this chapter is the optimization of D2D-based cell association using the notion of decoupled UL and DL association showcased in [23]. In specific, integer linear programming (ILP) optimization formulations are introduced in order to achieve efficient D2D cell association, aiming at minimizing the interference caused by D2D devices onto cellular links as well as improving the resource utilization of the network.

2.2 System Model

The main aim of this section is to present a number of D2D-aware cell association techniques and then dive into detailed analysis and comparison among each other. Complying with the LTE Release 12 specifications and the imminent integration of D2D in current and emerging networks, we follow a number of general design admitions according to up-to-date 3GPP standardization working documents [8]:

- D2D communication will operate with UL resources.
- The inband case of D2D communication is considered; in such case, the interference from a D2D user onto cellular connections (and vice versa) in a neighbouring cell could be substantial.
- The transmit power of D2D devices is defined and controlled by the serving cell (small or macro cell) based on fractional path-loss compensation power control, similar to cellular users [52]. In mathematical terms, the transmit power of a D2D user u that is associated with BS b is given by

$$P_t^{lu} = min\{P_{Max}, 10\log_{10}(M) + P_0 + \alpha PL_{blu}\}, \tag{2.1}$$

  where P_{Max} is the maximum transmit power of the device, M is the number of physical resource blocks (PRB) allocated to the device, P_0 is a normalized power value (in dBm), α is the path loss compensation factor and PL_{blu} is the path loss between the device u of link l and its serving cell b.

As already mentioned, D2D-aware CAS algorithms need to consider the fact that user devices might have subsequent cellular and direct (D2D) transmissions in subsequent subframes. Opposed to the current trend where a UE gets connected to the BS that provides the highest DL received power to it, the concept of UL/DL decoupling proposed in [23] has shown substantial gains by allowing users to choose different cells in the UL and DL. This idea is the basis of this chapter where it will be proven that the same concept is also applicable to D2D-related association. For the different cell association techniques to follow, interference minimization is the objective as it is one of the most critical challenges due to the ongoing cells' densification [14].

Next, we provide an ILP optimization framework based on the different association policies by considering the notion of decoupled association and the ability of the devices of a D2D pair to connect to different serving cells. Without loss of generality, unicast D2D links are assumed. The compared CAS methods are presented below.

- **Joint-Coupled (JC)**: The baseline case where devices of the same D2D pair are only allowed to connect to solely one cell (Joint). Furthermore, the D2D devices have the same UL and DL serving cell based on the DL received power (Coupled).

- **Joint-Decoupled (JD)**: The devices of a D2D pair connect to the same serving cell but the UL and DL associations are decoupled. In that method, the UL serving cell is the one that minimizes the UL interference to cellular communication (e.g. both devices can be associated with a macro BS in the DL and with a small cell BS in the UL).

- **Disjoint-Decoupled (DD)**: The communicating devices of a D2D pair are permitted to associate with different serving BSs; also, the two devices can have different UL and/or DL associations (i.e. separate serving BS per UL and DL session).

- **Hybrid-Decoupled (HD)**: In this case we combine both the Joint-Decoupled and the Disjoint-Decoupled cases to strike a balance between minimizing the interference and the resource usage.

In the last three cases D2D UEs are allowed to be associated with different serving cells in the UL (decoupled access) based on the minimum UL interference criterion. Before detailing the applied optimization framework, we need to define the following: the set of deployed BSs is denoted as $\mathcal{B}$ (including both macro and small cell BSs), the set of randomly distributed D2D links is $\mathcal{L}$, and lastly, U is the set of UEs that constitute the D2D links.

2.3 Problem Formulation

2.3.1 Joint-Coupled CAS

In this scheme we assume that both UEs of a D2D link are associated with the same BS according to DL received power calculations. This is the baseline method as it is applied in LTE networks. However, in that case the interference exerted by the D2D UEs can become harmful to the cellular transmissions, as exemplified in Fig. 2.1a. In this figure, *D2D 2* and *D2D 3*, both associated with the related macro BS can severely interfere with the proximate small cell UEs.

2.3.2 Joint-Decoupled CAS

This scheme's rationale is to associate both UEs of a D2D link with the BS that minimizes the link's UL interference. Fig. 2.1b represents this case. In this scenario, *D2D 3* is served in the UL by the small cell which results in the reduction of the transmit power of *D2D 3* as the pair is closer to the small cell. However, due to the joint association constraint, *D2D 2* is still associated to the macro cell.

Same as in [50], we extend the cell association optimization logic for D2D links, where the paired devices are both connected to the same serving BS [53].

For this reason, we define the following binary decision variable

$$y_{bl} = \begin{cases} 1, & \text{if D2D link } l \text{ is associated with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

Next, in order to view the problem of minimizing the interference coming from D2D UEs' transmissions, we need to define as $I_{bl} = mean\{I_{blu_1}, I_{blu_2}\}$ the mean value of the maximum interference generated by the two paired devices (u_1 and u_2 of link l) which are both coupled with BS b. The resulting interference formula for a D2D UE u of link l is $I_{blu} = max(P_t^{lu}\mathbf{G}_{\mathcal{B}'lu})$, where $\mathbf{G}_{\mathcal{B}'lu}$ is the matrix of link gains between user u and all BSs that belong to the set $\mathcal{B}' = \mathcal{B} - b$. P_t^{lu} accounts for the transmission power of the UE u of link l according to (3.18) and depends on its associated BS.

The interference-based optimization problem can be then formulated as follows

$$\min \sum_{b\in\mathcal{B}} \sum_{l\in\mathcal{L}} I_{bl} y_{bl} \tag{2.3}$$

subject to:

$$\sum_{b\in\mathcal{B}} y_{bl} = 1, \ \forall l \in \mathcal{L} \tag{2.3a}$$

$$\sum_{l\in\mathcal{L}} y_{bl} \leq K_b, \ \forall b \in \mathcal{B} \tag{2.3b}$$

$$y_{bl} \in \{0, 1\}, \ \forall b \in \mathcal{B}, l \in \mathcal{L} \tag{2.3c}$$

where constraint (2.3a) requires the sole association of a D2D link l to BS b, and (2.3b) provides an upper bound of the number of user links that can be associated with every BS b (K_b stands for the number of the available radio resources that can simultaneously serve the distributed user links per cell). The difference of this scheme compared to the Joint-Coupled baseline strategy is the decoupling of DL and UL for the D2D links located in the topology. Intuitively, but as also proven in the sequel, this method is very efficient in terms of resource utilization by blocking (utilizing) one RB only from its associated BS that controls the D2D transmission. On the other hand, this method lacks intelligence in terms of interference controllability as it associates both devices of a D2D link to one BS without giving the flexibility for separate association of the nodes that could be less harmful.

2.3.3 Disjoint-Decoupled CAS

In this scheme the paired devices can be also associated with different serving BSs as shown in Fig. 2.1c. This is expected to be an interference-restrictive strategy because every device connects to its closest serving BS. However, it is not efficient in terms of resource usage, simply because if both devices of a D2D pair are connected to two separate BSs, the resources used by these devices have to be allocated (*blocked*) for the D2D connection in both cells, contrary to the case where both devices are served by the same BS and the resources will be allocated (*blocked*) only in one cell. Hence, this scheme is interference optimal but it uses twice as much resources as the Joint schemes. In a similar fashion as before, we provide an optimization setting that aims to minimize the introduced interference caused by the D2D transmissions.

Firstly, we consider the following binary decision variable that indicates each UE's association with a BS

$$y_{blu} = \begin{cases} 1, & \text{if user } u \text{ of link } l \text{ associates with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

where $b \in \mathcal{B}$, $l \in \mathcal{L}$, and $u \in U$.

Thus, the interference minimization problem for the disjoint decoupled D2D cell association can be formulated as per below

$$\min \sum_{b\in\mathcal{B}}\sum_{l\in\mathcal{L}}\sum_{u\in U} I_{blu}y_{blu} \tag{2.5}$$

subject to:

$$\sum_{b\in\mathcal{B}} y_{blu} = 1,\ \forall l \in \mathcal{L},\ u \in U \tag{2.5a}$$

$$\sum_{l\in\mathcal{L}}\sum_{u\in U} y_{blu} \leq K_b,\ \forall b \in \mathcal{B} \tag{2.5b}$$

$$y_{blu} \in \{0,1\},\ \forall b \in \mathcal{B}, l \in \mathcal{L}, u \in U \tag{2.5c}$$

where I_{blu} is the maximum interference generated by a user device u of D2D link l if associated with BS b.

2.3.4 Hybrid-Decoupled CAS

In this case, we propose an interference-aware optimization problem with an objective to achieve resource usage efficiency. An effective and controllable resource utilization on top of an interference-aware method may well entail in balanced interference mitigation and resource efficiency impact. The Disjoint-Decoupled approach might be optimal in terms of interference but it is not efficient in terms of resource usage. On the other hand, the Joint-Decoupled approach is optimal in the sense of resource usage but lacks of satisfactory interference performance compared to the two methods mentioned above. Hence, the Hybrid-Decoupled problem tries to strike the balance between interference and resource utilization.

In order to realize this hybrid problem, an additional decision variable needs to be defined that will act as an indication of joint association for two devices that construct a D2D pair. This can be written as follows

$$z_{bl} = \begin{cases} 1, & \text{if link } l \text{ associates with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

Therefore, we propose a resource usage optimization problem that considers interference and formulate it as follows

$$\max \sum_{b \in \mathcal{B}} \sum_{l \in \mathcal{L}} z_{bl} \tag{2.7}$$

subject to:

$$\sum_{b \in \mathcal{B}} y_{blu} = 1, \ \forall l \in \mathcal{L}, \ u \in U \tag{2.7a}$$

$$\sum_{l \in \mathcal{L}} \sum_{u \in U} y_{blu} \leq K_b, \ \forall b \in \mathcal{B} \tag{2.7b}$$

$$\sum_{u \in U} I_{blu} z_{bl} \leq I_{th}, \ \forall b \in \mathcal{B}, l \in \mathcal{L} \tag{2.7c}$$

$$2 z_{bl} \leq \sum_{u} y_{blu}, \ \forall b \in \mathcal{B}, l \in \mathcal{L} \tag{2.7d}$$

$$\sum_{b \in \mathcal{B}} z_{bl} \leq 1, \ \forall l \in \mathcal{L} \tag{2.7e}$$

$$y_{blu}, z_{bl} \in \{0, 1\}, \ \forall b \in \mathcal{B}, l \in \mathcal{L}, u \in U \tag{2.7f}$$

As shown, the main objective is the maximization of the number of joint connections for the distributed D2D paired devices with respect to interference. Constraints (2.7a) and (2.7b) are defined as in problem (2.5). In (2.7c), a threshold that constrains the levels of interference if the devices of a link are jointly connected to a BS is added. This threshold can act as a weighting factor to decide if the focus of the algorithm should be interference (low I_{th}) or resource efficiency (high I_{th}). For this constraint, we limit the search to the n closest BSs to reduce the search space and consequently the complexity and size of the inequality matrix. Furthermore, constraint (2.7d) indicates that only if both devices of a link l will be associated with the same BS b, the value of z_{bl} variable equals to one (joint case). Lastly, (2.7e) stands for the restriction that each link's users can be associated with only one BS in the case of joint connection ($z_{bl} = 1$). Differently, they are disjointly connected to two separate BSs ($z_{bl} = 0$).

2.4 Simulation Setup

As deployment setup, the Vodafone LTE small cell testbed network deployment shown in Fig. 2.2 was used. The test network covers an area of approximately one square kilometre and includes two macro sites and 21 small cells represented by the black shapes and red dots, respectively. We use this existing testbed to simulate a relatively dense HetNet scenario. The propagation model is based on a high resolution 3D ray tracing path loss prediction model. This model takes into account clutter, terrain and building data and it guarantees a realistic and accurate propagation model. The user distribution is based on real traffic data extracted from the live network. We assume an inband operation of D2D where D2D UEs use the UL frequency band assigned for cellular (licensed) transmission. However, D2D and cellular UEs are scheduled on different resources which is termed as 'overlay' operation. The results are based on Monte Carlo simulations and are averaged over 100 simulation runs.

The operating frequency applied for the simulations is 2.6 GHz. The maximum transmit powers of macro-cells, small-cells and UEs are 46, 30 and 23 dBm respectively. The fractional path loss compensation power control algorithm in (3.18) is valued with $P_0 = -90$ dBm and $\alpha = 0.8$ (considered as an optimal value in [54]. An average number of links of 336 is herein considered. Further, without loss of generality, I_{th} is set to -130 dB as it is proven that this value creates multiple instances of association ambiguity and is worth investigating. The

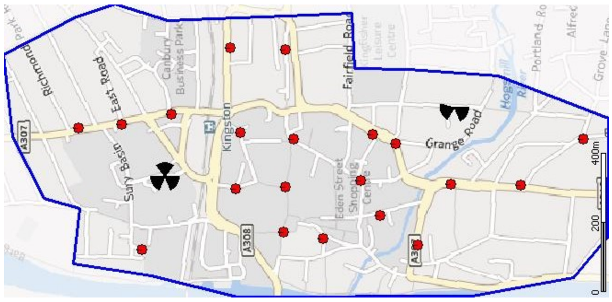


Figure 2.2: Vodafone Small cell enabled LTE test network.

reason why the interference comparisons have been made in dB is to scale the estimated interference measurements (resulting from the multiplication of the interfering transmitter's power with the corresponding link gain) down to a smaller and more conceivable number.

The next section features a set of results evaluating the proposed cell association methodologies proposed in section 2.3. Finally, for ease of comprehension, we assume that each D2D pair is allocated with one RB per BS.

2.5 Performance Evaluation

In this section, a set of results is presented to evaluate the proposed CAS optimization techniques. To produce the results to follow, Vodafone's radio planning tool, namely ATOLL[1], provided us with realistic system parameters (i.e. realistic user locations, precise power, SINR and path-loss estimations) that were used as input to our optimization simulator, run in MATLAB. For the produced results, MATLAB's *intlinprog* function has been leveraged to output the optimization outcomes. First, Fig. 2.3 showcases the mean UL interference exerted by the D2D UEs onto cellular transmissions in relation to varying D2D link length. The interference values are normalized relative to the DD case in order to highlight the different interference levels compared to this interference optimal scheme. The JC

[1]ATOLL is an open and flexible multi-technology RF platform supporting GSM/GPRS/EDGE, UMTS/HSPA, LTE/LTE-Advanced, Wi-Fi, WiMAX, Microwave and so forth.

and JD schemes show an increasing interference trend with the link length where the interference levels are around 3 dB (twice) and almost 6 dB (4 times) more than the DD scheme at 100 m and 150 m link length, respectively. This happens because the more the link range increases the more suboptimal the joint association schemes are as forcing distant devices to associate with the same BS entails a higher transmit power of these devices and a higher interference to neighbouring cells. The HD method provides a trade-off between the Joint and DD schemes as it retains an almost constant interference level that is around 1 dB higher than the DD scheme.

As mentioned before, if a D2D pair is served by one BS it is assumed to use only one RB over the whole network as this radio resource is allocated for this D2D pair in this BS only. However, if the devices of a pair are associated to different BSs then it is assumed that this pair is blocking two RBs over the network since one RB has to be reserved for that pair in both BSs. Fig. 2.4 shows the average D2D resource usage per BS in relation to the D2D link length. In this figure, a constant resource usage for the JC and JD association schemes is shown. This trend can be explained by the fact that the D2D pairs are jointly associated to the same BS regardless of the link length. Hence, each D2D pair uses one RB independent of the link length. However, the DD scheme shows an increasing RB usage with the link length. This happens because the probability of disjoint association increases proportionally with the link length and so does the D2D resource usage in the whole network, since the disjoint D2D pair uses twice as much RBs as the joint case. Similar as before, the HD scheme offers a compromise between the joint and disjoint schemes as the main target of the optimization problem is to ameliorate the resource usage efficiency with respect to the exerted interference. The HD method achieves a reduction of resource usage of about 45% at 150 m link length compared to the DD method. Therefore, it can be noted that the HD scheme effectively strikes a balance between the UL

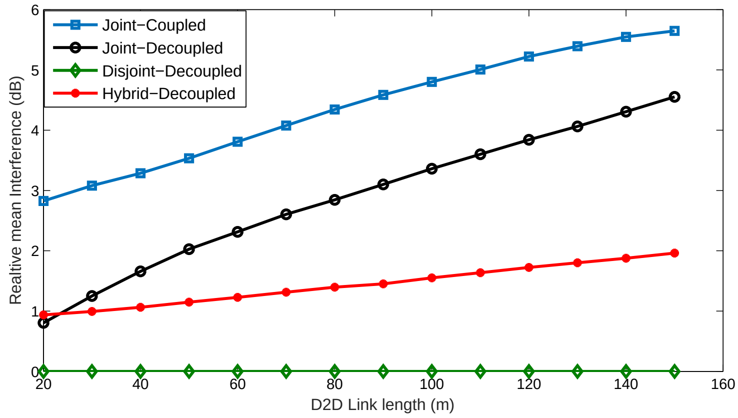


Figure 2.3: Mean UL interference from D2D devices onto cellular transmissions.

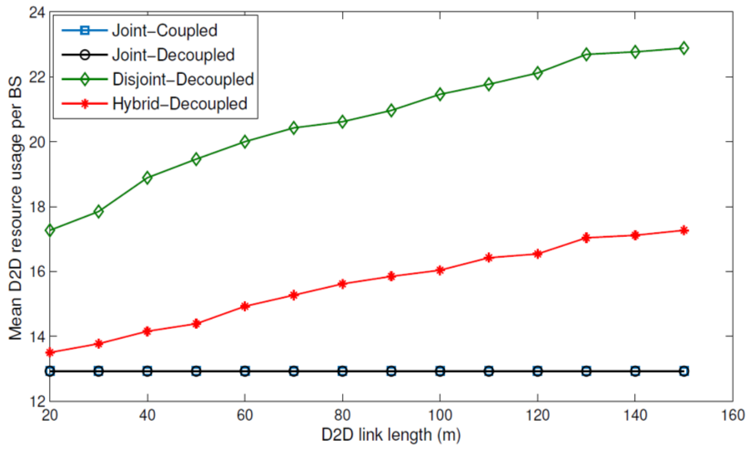


Figure 2.4: Mean resource utilization for D2D per base station.

interference and resource efficiency. An important thing is that it is tunable and it can also be controlled by setting the I_{th} according to the operator's needs.

Finally, the cumulative distribution function (CDF) of the D2D UEs transmit power is depicted in Fig. 2.5. It shows that the JC and DD schemes have the highest and lowest transmit power distributions, respectively, with a difference of more than 5 dB at 50% of the CDF. The gap in the CDF increases the higher the transmit power is. The HD method's distribution lies in between the JC and JD distributions and gets closer to the DD the higher the transmit power gets. This proves that the HD scheme can result in a reduction of transmit power that varies between 3-5 dBs which is deemed crucial for the battery of the user devices.

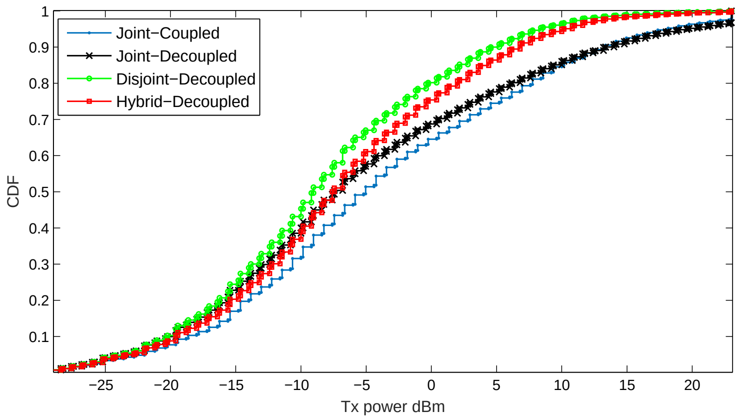


Figure 2.5: CDF of the devices' transmit power.

Chapter 3

Cell Association Optimization for D2D Underlaying Fractional Frequency Reuse based Cellular Networks

3.1 Introduction

In this section, the focus is turned on the inband underlay property of D2D communications where the paired users can reuse the available cellular (licensed) spectrum. As already mentioned, D2D links that underlay a cellular infrastructure will be mainly controlled by the network to ensure higher spectrum controllability [55]. In terms of network operation, D2D communications provide some new challenges due to their dynamic nature and the new intra/inter cell interference patterns that will generate. The coexistence of conventional cellular links and D2D pairs perplexes the problem of resource allocation due to the limited physical resources. Each BS can serve only a limited number of connections simultaneously, and for this reason, the overall traffic load should be efficiently balanced among the serving BSs. Therefore, it is important to apply efficient D2D-based cell association techniques in order to not only provide satisfactory QoS for all UEs, but also enhance the spectral efficiency and system's capacity to accommodate more users to serve.

Due to the ongoing proliferation of social networking based applications, the chance for two users in close proximity to share data between each other might significantly increase in the future, leading to the irregular emergence of multiple D2D pairs in future networks that need to be orchestrated. In the following subsection, a statistical based characterization to examine the possible emergence of D2D links where part of the cell-edge ones might cross different cell coverage areas is provided. The reason for examining the latter is the ambiguity of a crossing link to be associated only with one of the candidate serving BSs. Like in conventional cellular communications, several aspects might influence the cell-edge D2D links performance and its association with a specific BS. These are, for example, the link range, path loss characterization and the distance from each BS.

3.1.1 Statistical bound for macro-cell crossing D2D links

We consider a highly dense scenario where multiple D2D links are uniformly distributed in a seven-hexagonal cell scenario (the establishment procedure of D2Ds is out of the scope of our contributions, however a summary of the proximity-based session initiation ([56]) and link formation steps can be found in [11][13]). Part of them will be consisting of links where the involved DUEs are expected to fall within the geographical area of different BSs (red links shown in Figure 3.1). As we are now witnessing further cell densification and overall decrease of the cell size in order to increase spatial capacity of future networks, the case of two nodes being located in different cells might become a notable proportion of the D2D communication links.

We correlate the problem of having cell-crossing D2D links with the Buffon's Needle (BN) problem [57]. This problem examines the probability that a needle lies in a position where it intersects one of the parallel lines when dropped on a ruled two-dimensional space. In the D2D case, this probability could refer to a D2D link where the two DUEs geographically belong to different cell. Below, we provide an estimation of the lower bound regarding the number of D2D pairs for which it is not straightforward which BS will undertake their control. In a heterogeneous network setting, this number can be distinctly higher due to the existence of small cells within the serving area of macro BSs, however this is out of scope in this chapter and could be a future subject of research.

Let us assume hexagonal cells with dimension $h = 2d$ (Figure 3.1) and define with l_{n_1,n_2} the distance between two DUEs n_1 and n_2, it is proven that the probability p_0 that both UEs are located in the same cell is approximated as follows [57]:

$$p_0 = 1 - \frac{1}{3}\left(\frac{l_{n_1,n_2}}{h}\right)^2 - \frac{l_{n_1,n_2}}{h}\left(4 - \sqrt{3}\frac{l_{n_1,n_2}}{h}\right)\frac{1}{\pi}. \tag{3.1}$$

Consequently, the crossing probability for D2D links is the complement of (3.1):

$$p = 1 - p_0 = \frac{1}{3}\left(\frac{l_{n_1,n_2}}{h}\right)^2 + \frac{l_{n_1,n_2}}{h}\left(4 - \sqrt{3}\frac{l_{n_1,n_2}}{h}\right)\frac{1}{\pi}. \tag{3.2}$$

In the case of a highly congested network, this probability could provide a statistical approximation of the ratio of crossing D2D links in relation to the total number of D2D links in the network. We denote this *crossing ratio* with CR, thus $CR = \frac{N_{cross}}{N_{total}}$. To visualize this, we randomly distributed a large designated number of D2D pairs in a hexagonal seven-cell scenario and compared the previously defined probability with our implementation's ratio (Figure 3.2). By running extensive Monte Carlo simulations (1000 iterations) and for fixed number of links, we observe that the output (i.e. CR) becomes approximately tangent to

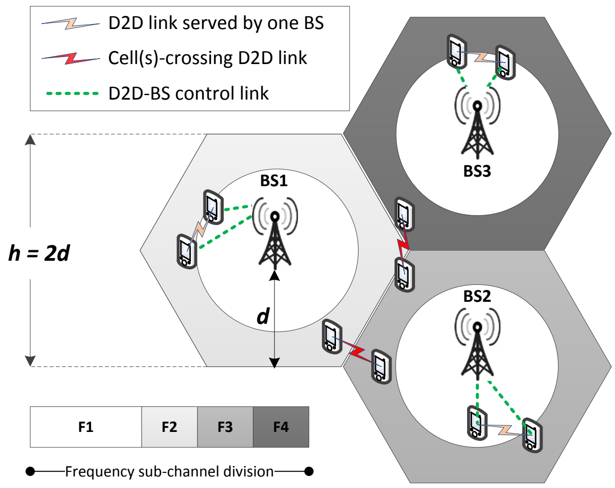


Figure 3.1: D2D communication links in a multi-cell environment (with $h = 2d$). A number of links might cross the boundaries between neighbouring cells, hence link nodes could be conventionally connected to different BSs.

the BN problem's probability estimations. In this figure, the depicted probabilities and ratios are calculated for varying cell dimensions and for different link ranges. As shown, the link crossing probability increases in proportion with the range of the D2D communication link and decreases for cells with larger radius. In real world scenarios where the coverage areas of the deployed BSs are irregular, the percentage of crossing D2D links is expected to be significantly higher in some cases. Also, it is worth pointing out that as we are now witnessing further cell densification and overall decrease of the cell size in order to increase spatial capacity of emerging and future networks, the case of two nodes being located in different cells might become a notable proportion of the D2D communication links.

However, the above analysis corresponds to an ideal scenario where cells form hexagonal shapes of the same size and are served by center-located BSs. To reflect a more realistic deployment scenario, BSs can be distributed according to a Poisson Point Process (PPP) in space, and each one of them controls a

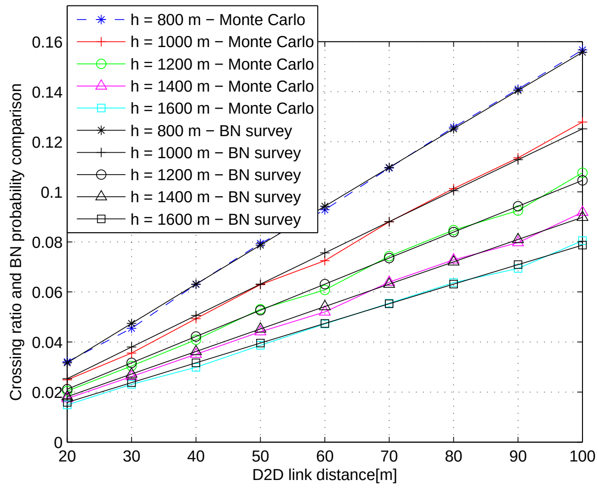


Figure 3.2: Buffon's Needle survey and Monte Carlo simulations comparison. This figure signifies the probability of having a D2D link crossing the boundaries of neighbouring cells. Convergence is achieved for a high number of iterations.

Voronoi region (cell[1]) with a random area as also contemplated in [58]. Because of the irregularity of the cell shapes as well as the PPP distribution of the BSs, the terms inner (interior) and outer (cell-edge) user do not have the same geometrical interpretation as in hexagonal layouts. Hence, in order for a BS to characterize a user as an interior or a cell-edge one, a pre-specified SINR threshold is defined and compared with a user's average SINR; when it is below this threshold, it is labeled as a cell-edge UE, otherwise as interior.

3.1.2 Closely related work

Cell association in cellular networks has been a critical challenge for network operators due to the ongoing increase of user demands in an underlying resource-limited infrastructure. The scope of optimizing cell association is to enhance network capacity and accommodate more users simultaneously with respect to their QoS requirements. Although it is a well-investigated area of research, cell association considering the integration of D2D communications as an underlay in

[1]Geometrically, the cell, serving a set of supported users, is determined by a closest-BS association, thus the mobiles are the ones to compose the Voronoi partition/tesselation in space

cellular networks is a rather unexplored and is the main difference of our work hereafter compared to the existing ones.

As it was also discussed in the previous chapter, an important question that arises is whether the users that form a D2D pair will be jointly associated with a single BS or with two separate ones. In the former, the signaling burden is alleviated because both nodes that constitute a D2D pair are associated to the same BS, whereas in the latter extra BS signaling exchange is needed and communication latency increases. In this chapter, we only consider the first case where both users of a D2D link will be coupled with the same BS in the underlay case [53], according to a set of criteria that will be mentioned in the sequel.

It has to noted that relevant literature which considers explicitly the existence of cell-crossing D2D links is quite limited. Significant works that take into account inter-cell D2D links are [59] and [60], both elaborating on the issue of radio resource allocation. The former focuses on optimizing the achievable aggregate network throughput in a three-cell scenario where the D2D users are eligible to reuse the downlink cellular resources. The latter proposes a game-theoretical model where the BSs are competing resources for serving the D2D-related demands, and proceeds by devising a resource allocation algorithm based on Nash equilibrium derivations. In other significant works, cell-crossing links' emergence is also implied as inter-cell D2D UEs, eligible to connect, might be scattered via PPP modeling in two different cells [61][62].

3.1.3 Contribution and Structure

The key aim here is to provide novel cell association optimization solutions for varying network congestion episodes that will boost network capacity to accommodate increased number of users simultaneously. The different proposed solutions belong to an overarching optimization framework which can be deemed as a toolkit for a network operator to optimize network performance based on different selected criteria. The underlay concept of D2D communications (that entails the most spectrum-efficient property among all), combined with an effective balancing of the D2D links along the network, can lead to valuable resource savings. Also, resource allocation for D2D communications needs to be designed in a way that network throughput is boosted. To this direction, and due to the NP-hardness of the joint cell association and resource allocation, the problem is tackled by decoupling it into two sub-problems: first, the cell association problem that can be solved via ILP tools and, second, the resource allocation which can be efficiently addressed by an inherently randomized RA algorithm with low computational complexity. In that case, the output solution of the selected cell association optimization problem will become the input for the resource allocation (RA) technique that will be provided. Thus, a linear time resource allocation algorithm on top of an optimized CAS configuration is introduced to offer substantial network throughput performance other than resource efficiency.

3.2 System Model

3.2.1 Signalling overhead

The investigated D2D communication links can be categorized into two groups: inner and outer links. The first one includes those D2D links that their paired nodes are both located in a single cell, far from its edges, and are being controlled by the same BS. The second group corresponds to those links that are located in the edge of a cell and their nodes might belong to two different serving geographical areas, as mentioned before. Considering the latter, we presume that the two DUEs could be associated with different BSs. However, this will eventually increase latency and signaling overhead due to the need of the two involved BSs to coordinate the communication (Figure 3.3a). Thus, in order to limit these impeding factors, we assume that the overall signaling overhead for

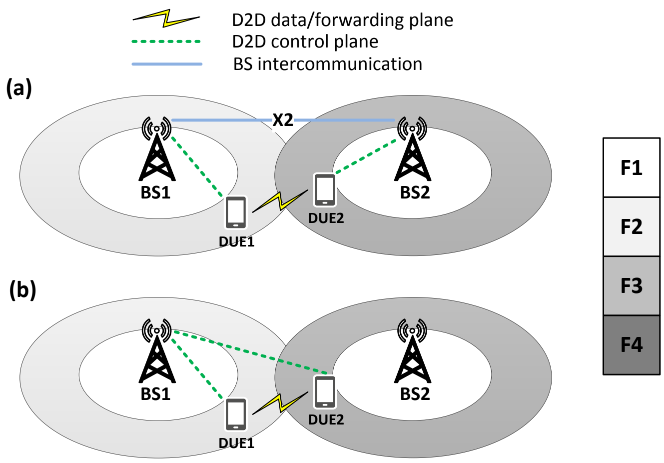


Figure 3.3: Signalling exchange for the control of a cell-edge D2D pair: (a) Disjoint DUE association, (b) Joint DUE association (no exchange).

this case study can be reduced by providing explicit association of each pair to a single BS, aiming to avoid any BS intercommunication to exchange information (Figure 3.3b) [53]. How to properly select a specific BS for associating with a D2D pair will be explained in the sequel. This approach corresponds to a more realistic and trustworthy networking system as it is easier to be implemented compared to multi-BS association of the connected users. Let us now consider that the number of cell-edge D2D pairs is *at least* N_{cross}. This signaling reduction implies pairwise association with a single BS for each formed D2D pair and, therefore, a signaling exchange saving of $\frac{N_{cross}}{2}$ cooperative flows among BSs can be achieved.

We assume that the cell association problem that jointly couples both DUEs of each D2D pair with the same cell/BS can be solved by a central entity that firstly accumulates all of the users' location coordinates and then processes the input to produce the requested joint cell association pattern for D2D users. Therefore, for the part of the problem that is investigated, no BS intercommunication is required because both DUEs will jointly associate with solely one BS. The decision of which BSs are the candidate ones for a D2D pair to associate with depends on the mean path-loss estimations (distance based) and is analyzed in the sequel.

3.2.2 Fractional Frequency Reuse (FFR)

Our approach is based on the consideration of FFR as an interference-limiting method. In [63], enhanced soft FFR is considered as a promising key technology to achieve large-scale cooperative radio resource management (LS-CRRM) in future 5G networks due to its interference controllability attribute. Motivated by it, we further consider the integration of D2D communications and apply a differentiated FFR scheme where D2D users will be assigned resources from a RB pool that its content depends on the DUEs' location and the respective BSs to serve them [64]. A frequency reuse factor (FRF) of three is used for the cell-edge (outer) areas, as depicted in Figure 3.1. With this modelling, CUEs that are located in the inner cell areas can use part of the whole frequency band (i.e. *F1*), whereas outer cell CUEs can use one third of the remainder (*F2*, *F3* or *F4*). The available RB pool for users located in inner region (N_{Inner}) is proportional to the interior-area radius and is twice the size of the pool corresponding to the cell-edge users (N_{Outer}) [65]. However, the differentiation of this FFR scheme concerns the D2D communication links. Compared to the conventional soft FFR that is applied for CUEs, if D2D UEs are located in cell inner region, they can utilize resources from the frequency sub-bands that cellular users do not use within the same cell (e.g. in Figure 3.1, a D2D link located in the inner region of cell 3 can be assigned resources only from sub-bands *F2* and *F3*). On the other hand, if DUEs are located in a cell's outer area, they can utilize resources from all available spectrum except for the sub-band that can be exploited by cellular users in identical cell outer area (again, for a cell-edge D2D link in cell 3, the sub-bands *F1*, *F2* and *F3* would compose its available RB pool).

With this interference-aware method, inner-region D2D and cellular transmissions are happening in orthogonal channels. However, intra-cell interference still exists but can only be exerted from outer-cell DUEs and inner-cell CUEs (and vice versa) or multiple DUEs that might utilize the same resources. Regarding inter-cell interference, outer-cell D2D links can experience interference by adjacent outer-cell CUEs.

3.2.3 Basic notations and definitions

As also discussed in the previous subchapter, the set of BSs is denoted with $\mathcal{B} = \{b_1, b_2, \ldots, b_{|\mathcal{B}|}\}$, whereas the set of D2D links is $\mathcal{L} = \{l_1, l_2, \ldots, l_{|\mathcal{L}|}\}$ (randomly distributed in a hexagonal multi-cell topology). The $|\cdot|$ notation declares the cardinality of a set. All BSs have the same number of resources K_b. However, the available RB pool for each $b \in \mathcal{B}$ is different and depends on the discussed FFR scheme, as shown in Table 3.1. In the context of this work, each association of a D2D link with a BS implies the occupation of a single RB. As a consequence, the total number of D2D associations with a specific BS will be equal to the number of RBs allocated by the same BS.

Let us further define by c_{lb} the *cost* of a D2D link l connected to BS b; this can be considered as the average path-loss (distance-based) of connecting both nodes n_1 and n_2 of a D2D pair at the same BS and is estimated as follows:

$$c_{lb} = \frac{PL_{n_1,b} + PL_{n_2,b}}{2} \tag{3.3}$$

where $PL_{n_i,b} = 128.1 + 37.6 \log_{10} r_{n_i,b}$ is the path loss (in dB) between BS b and DUE n_i, for $i = 1, 2$. In the previous formula, $r_{n_i,b}$ is the DUE-BS distance (in kilometres). For the estimation of c_{lb}, $PL_{n_i,b}$ values are converted from dB values to ratios. For current and emerging cellular networks, where connected DUEs might have subsequent direct (D2D) and cellular UL/DL transmissions, this cost metric represents the need to stay "as close as possible" to the serving BS to support both communication types that can happen in short and sequential time

Table 3.1: Resource availability per cell according to differentiated FFR

Cell id (b)	**Inner-region D2D**	**Outer-region D2D**
$b = 1$	$\{F_3\} \cup \{F_4\}$	$\{F_1\} \cup \{F_3\} \cup \{F_4\}$
$b = 2 \cdot i^*$	$\{F_2\} \cup \{F_4\}$	$\{F_1\} \cup \{F_2\} \cup \{F_4\}$
$b = 2 \cdot i + 1$	$\{F_2\} \cup \{F_3\}$	$\{F_1\} \cup \{F_2\} \cup \{F_3\}$

$^* i = 1, 2, 3, \ldots$

epochs (the lower the value of c_{lb} the bigger the probability to associate with the closest BS). Furthermore, because in this work the focus is turned on the UL, a user's association with a BS should be preferably decided by its estimated path loss to it and not by the traditional downlink received signal-based criterion in cellular networks [24]. To this end, and based on the aforementioned analysis for ensuring reduced signalling overhead, the D2D links that are characterized by association ambiguity (i.e. two nodes should be normally associated with different BSs) are coupled with the BS that achieves the minimum average path loss for each pair of nodes.

In order to formulate the problem of the D2D cell association and formulate it mathematically, the following binary variable needs to be defined:

$$y_{lb} = \begin{cases} 1, & \text{if link } l \text{ is connected to BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

The sequence of the y_{lb} values will construct a vector that defines the solution of the ILP optimization settings that will follow. This vector can be represented as:

$$\mathbf{y} = \left[y_{11}, y_{21}, \ldots, y_{|\mathcal{L}|1}, \ldots, y_{1|\mathcal{B}|}, \ldots, y_{|\mathcal{L}||\mathcal{B}|} \right]^{\intercal}. \tag{3.5}$$

It is clear that only $|\mathcal{L}|$ values of it can equal to one due to the sole association of a D2D link with only one BS.

3.2.4 Joint cell association and resource allocation

In this paragraph, the joint cell association and resource allocation problem is introduced (problem (3.6)). As it will be further analyzed in the sequel, the first objective corresponds to the need for balancing the number of connections in order to achieve a resource efficient orchestration and constitutes the cell association part of the problem. On the other hand, D2D sum rate maximization is introduced as the second objective and aims at optimizing the resource allocation for D2D communications.

In continuity of the definitions presented above, we further denote with R_{lbk} the achievable throughput for link l that associates with BS b and utilizes the RB k, and also define x_{lbk} as a binary decision variable that indicates whether the link l, associated with BS b, is assigned with RB k or not. Then, the joint problem can be formulated as follows:

$$\min \left\{ \sum_{b\in\mathcal{B}} \left(\sum_{l\in\mathcal{L}} y_{lb} \right)^2 \; ; \; -\sum_{l\in\mathcal{L}} \sum_{b\in\mathcal{B}} \sum_{k\in\mathcal{K}} R_{lbk} x_{lbk} \right\} \tag{3.6}$$

$$\text{s.t.} \; \sum_{b\in\mathcal{B}_l} y_{lb} = 1, \; \forall l \in \mathcal{L} \tag{3.6a}$$

$$\sum_{l\in\mathcal{L}} y_{lb} \leq K_b, \; \forall b \in \mathcal{B} \tag{3.6b}$$

$$\sum_{b\in\mathcal{B}_l} c_{lb} y_{lb} \leq c_{\text{th}}, \; \forall l \in \mathcal{L} \tag{3.6c}$$

$$\sum_{b\in\mathcal{B}_l} \sum_{k\in\mathcal{K}} R_{lbk} x_{lbk} \geq R_{\text{th}}, \; \forall l \in \mathcal{L} \tag{3.6d}$$

$$\sum_{b\in\mathcal{B}_l} \sum_{k\in\mathcal{K}} x_{lbk} = 1, \; \forall l \in \mathcal{L} \tag{3.6e}$$

$$\sum_{k\in\mathcal{K}_b} x_{lbk} = y_{lb}, \; \forall l \in \mathcal{L}, \; \forall b \in \mathcal{B} \tag{3.6f}$$

$$y_{lb}, x_{lbk} \in \{0, 1\}, \; \forall l \in \mathcal{L}, \; \forall b \in \mathcal{B}, \; \forall k \in \mathcal{K}, \tag{3.6g}$$

where $\mathcal{B}_l$ is the set of candidate BSs to be associated with the link l. Each

D2D pair cannot be associated with whichever BS. Constraints (3.6a) require that each link l will be associated with only one of the BSs that belongs to the $\mathcal{B}_l$ set. Inequality constraints (3.6b) are introduced to avoid any resource availability violation for each BS b. (3.6c) accounts for the cost values (eq. (3.3)) not to be above a pre-defined cost threshold c_{th}. Constraint (3.6d) accounts for satisfying each D2D link's rate threshold, whereas (3.6e) means that each D2D link l can be associated with only one BS and be assigned with only one RB. Then, (3.6f) signifies that a link l that is associated with a BS b can be only assigned with a RB k that stems from the BS's b available resource pool (i.e. $\mathcal{K}_b$). Finally, (3.6g) ensures the binary assignment of the $\mathbf{y}$ and $\mathbf{x}$ vector's values.

Due to the reusability of a RB by potentially more than one D2D pair as well as the existence of multiple D2D links, this problem falls into the nature of MINLP[2] optimization problems that are hard to be solved in polynomial time and optimal solution cannot be acquired unless a number of constraints' relaxation applies. It is also worth pointing out that cell association and RB allocation take place in time scales that can differ multiple orders of magnitude and therefore, looking at this problem at the time domain, it can be concluded that in real-world applications it is a natural approach to decompose the problem as presented in later stage. To this end, and in order to reduce the complexity and hardness, the decoupling of the joint problem into two sub-problems is proposed; first, we solve the ILP cell association problem, and then, following the produced D2D association pattern, we apply a low-complexity resource allocation heuristic algorithm.

[2]MINLP refers to optimization problems with continuous and discrete variables and non-linear functions in the objective function and/or the constraints. In problem (3.6), the first objective falls into the case of non-linearity.

3.3 D2D Cell Association Problem Description

A set of cell association-based optimization problems for DUEs will be herein presented. The basic idea is to introduce an optimal framework for D2D links that considers spectrum efficiency as well as interference restriction in a multi-cell network. To this end, in the following paragraphs we proceed with proposing a multi-objective cell association optimization framework, namely *MOCA*, that consists of a number of different D2D cell association formulations with different objective functions. It is anticipated that, according to varying network traffic scenarios, this set of optimization problems can be considered as an add-on feature for the network operator to be able to choose among the different association policies.

3.3.1 Resource-aware Cell Association Optimization: MOCA-I

1. *Motivation*: Cell association is highly correlated with the ability of the network infrastructure to accommodate a significant number of connections simultaneously. However, the integration of D2D paradigm in emerging wireless systems urges network operators to contemplate how the DUEs' association problem should be addressed in order to efficiently exploit its resource reuse ability and, thus, avoid any resource blocking by dedicating part of the spectrum for it (overlaid D2D communications). On top of this, interference exerted to (from) the cellular communications from (to) D2D links and among multiple D2D transmissions needs to be taken into consideration. This means that, if the limited available resources provided by a macro BS are potentially over-utilized by multiple D2D links in a cell, this can bring in undesired interference patterns. For this reason, D2D user association should attain a balanced D2D-based link orchestration with

respect to resource efficiency.

2. *Problem formulation*: This problem can be mathematically formulated as follows:

$$\min \sum_{b\in\mathcal{B}} \left(\sum_{l\in\mathcal{L}} y_{lb} \right)^2 \tag{3.7}$$

subject to:

$$\sum_{b\in\mathcal{B}_l} y_{lb} = 1, \ \forall l \in \mathcal{L} \tag{3.7a}$$

$$\sum_{l\in\mathcal{L}} y_{lb} \leq K_b, \ \forall b \in \mathcal{B} \tag{3.7b}$$

$$\sum_{b\in\mathcal{B}_l} c_{lb} y_{lb} \leq c_{\mathrm{th}}, \ \forall l \in \mathcal{L} \tag{3.7c}$$

$$y_{lb} \in \{0,1\}, \ \forall l \in \mathcal{L}, \ \forall b \in \mathcal{B}. \tag{3.7d}$$

It is a common practice to assume that each link can only be associated with a small number of BSs considering its location. Constraint (3.7a) requires that each link l will be associated with only one of these BSs that belong to the $\mathcal{B}_l$ set. Inequality constraint (3.7b) is introduced to avoid any resource availability violation for each BS b. (3.7c) accounts for the cost values to be below a pre-defined cost threshold c_{th}. Finally, (3.7d) ensures the binary assignment of the $\mathbf{y}$ vector's values.

The above constraints were detailed in 3.2.4. We note that the above is in essence a non-linear integer optimization problem which is not suitable to be solved via powerful available toolboxes on integer linear mathematical programming. This problem can be then re-formulated as an integer linear program if viewed as a max-min optimization problem in the following way:

$$\max \quad z \tag{3.8}$$

subject to:

$$z \leq \sum_{l \in \mathcal{L}} y_{lb}, \ \forall b \in \mathcal{B} \tag{3.8a}$$

$$\sum_{b \in \mathcal{B}_l} y_{lb} = 1, \ \forall l \in \mathcal{L} \tag{3.8b}$$

$$\sum_{l \in \mathcal{L}} y_{lb} \leq K_b, \ \forall b \in \mathcal{B} \tag{3.8c}$$

$$\sum_{b \in \mathcal{B}_l} c_{lb} y_{lb} \leq c_{\mathrm{th}}, \ \forall l \in \mathcal{L} \tag{3.8d}$$

$$y_{lb} \in \{0, 1\}, \ \forall l \in \mathcal{L}, \ \forall b \in \mathcal{B}. \tag{3.8e}$$

z is a positive number and the maximization of it accounts for achieving as much as possible a balance in terms of number of links associated with the deployed BSs. Constraint (3.8a) is also added to achieve the desired balance of D2D links' association within the multi-cell topology by keeping the sums of the second part of the inequality close to the value of z. This optimization problem is an ILP problem that due to the unimodular property of its aggregate inequality matrix (i.e. determinant of every square inequality sub-matrix equals to 1) can be solved efficiently and easily, since it resembles the computational complexity of the corresponding linear (fractional) program. Additionally, let us denote by $\mathbf{s}$ the final solution vector for this optimization problem. It consists of the decision variable $\mathbf{y}$ as well as the z-constrained integer variable and can be mathematically represented as $\mathbf{s}$ $= [\mathbf{y}; z]$.

3.3.2 Joint connectivity cost & RB reuse optimization: MOCA-II

1. *Motivation*: As already mentioned, with the advent of the data-driven era and the ongoing user densification, the probability of two users to communicate directly increases. Especially in mass events, such as concerts or football games, where the incoming data requests are highly correlated to the event, the need to support multiple local communications arises. However, due to the limited channel resources, and in order to support these multiple connections, some of the available resources might be reused by multiple links within a cell. This can be translated to interference effects among the users that use the same RBs. To this end, an optimization problem is proposed as pertain to the issue of resource optimization usage and, more specifically, to efficiently minimize the overall RB reuse levels for highly congested scenarios. We herein provide D2D-BS association, in parallel with RB utilization awareness, so that the reuse rate of RBs (which are assigned based on the differentiated FFR method) is potentially minimized.

2. *Problem formulation*: We formulate the aforementioned problem as a bi-objective optimization setting. First, we define the following decision variable:

$$\tau_{rb} = \begin{cases} 1, & \text{if RB } r \text{ of BS } b \text{ is used} \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

Additionally, we denote with ρ_{rb} an index that captures how many times a RB r is assigned by the BS b. Based on the above definitions, we formulate the following optimization problem which provides optimal D2D cell association with the prospect of minimizing the reuse of RBs in the network:

$$\min \left\{ \sum_{l\in\mathcal{L}} \sum_{b\in\mathcal{B}} c_{lb} y_{lb} \ ; \ \sum_{r\in\mathcal{R}} \sum_{b\in\mathcal{B}} \rho_{rb} \tau_{rb} \right\} \tag{3.10}$$

subject to:

$$\sum_{b\in\mathcal{B}_l} y_{lb} = 1, \ \forall l \in \mathcal{L} \tag{3.10a}$$

$$\sum_{l\in\mathcal{L}} y_{lb} \leq K_b, \ \forall b \in \mathcal{B} \tag{3.10b}$$

$$\sum_{r\in\mathcal{R}} \tau_{rb} \leq K_b, \ \forall b \in \mathcal{B} \tag{3.10c}$$

$$\sum_{r\in\mathcal{R}} \tau_{rb} \leq \sum_{l\in\mathcal{L}} y_{lb}, \ \forall b \in \mathcal{B} \tag{3.10d}$$

$$\sum_{b\in\mathcal{B}_l} c_{lb} y_{lb} \leq c_{\text{th}}, \ \forall l \in \mathcal{L} \tag{3.10e}$$

$$y_{lb}, \tau_{rb} \in \{0, 1\}, \ \forall l \in \mathcal{L}, \ \forall b \in \mathcal{B}, \ \forall r \in \mathcal{R}. \tag{3.10f}$$

Following the notations used above, $\mathcal{R}$ accounts for the set of total available resources (e.g. for a 10 MHz LTE-based system bandwidth, set $\mathcal{R}$ contains 50 physical RBs in total, according to 3GPP specifications). However, according to the principles of the differentiated FFR scheme, each BS b provides a subset of the total available resources (Table 3.1). According to it, each inner-region D2D pair can be assigned a channel resource out of 20 RBs, whereas for the outer-region ones the number is doubled. Hence, the resources' upper bound per cell is 40. Constraints (3.10a), (3.10b) were described in the previous subsection. Further, we note that constraints (3.10c) are logically redundant since they are implied by the constraints in (3.10d), but we include them in the formulation in order to reduce the search effort and runtime (i.e., reducing further the search space).

Additionally, it has to be highlighted that the proposed formulation can be used in other frequency reuse techniques (or even different frequency reuse factors) that mainly aim to address the inter-cell interference coordination problem in multi-cell networks.

3.3.3 Joint interference-aware & resource-aware cell association optimization: MOCA-III

1. *Motivation*: A very important issue that needs to be addressed is the potential interference developed due to the co-existence of multiple D2D pairs and CUEs. According to the applied FFR technique, a cell-edge or crossing D2D pair that associates with a BS can mainly cause interference to cellular transmission of adjacent outer-cell regions that might utilize the same resource. Similarly, during the UL session that D2D communication is expected to happen, the D2D receiver suffers interference from the CUE that transmits to its coupled BS. Because of the limited resources for outer cellular UEs, the probability that a CUE will utilize the same resource with a neighbouring DUE becomes high. Considering that a CUE has a specified interference range, the possibility that a D2D link will be harmed needs to be avoided. An example is given in Figure 3.4.

   Especially in highly dense scenarios, the need to avoid immense interference is of paramount importance. By applying a joint optimization framework that regards not only the resource availability but also the existence of potential cellular interferers, this can entail better achievable system throughput and overall performance improvement in the long run.

2. *Problem formulation*: In order to mathematically formulate this problem, we introduce a penalty factor ϑ_{lb} (natural number) to represent the number of cellular users that are candidate interferers to a D2D link l if it is associated with BS b. Each ϑ_{lb} value accounts for the number of neighbouring cells' potential cellular interferers in an arithmetic/statistical fashion (e.g.

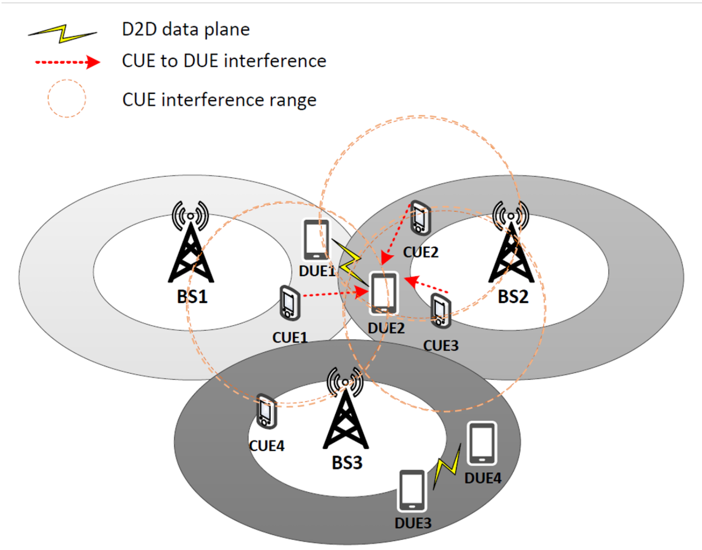


Figure 3.4: CUEs to DUEs interference depiction. Possible association of the D2D link with BS1 will give higher probability of interfering with a cellular user that is located in the serving area of BS2.

in Figure 3.4, if the intersecting link associates with BS1, then $\vartheta_{lb} = 2$, for $b = 1$). For ease of comprehension, we first formulate the problem as a solely interference-aware minimization problem as follows:

$$\min \sum_{l \in \mathcal{L}} \sum_{b \in \mathcal{B}} \vartheta_{lb} y_{lb} \tag{3.11}$$

subject to:

$$\sum_{b \in \mathcal{B}_l} y_{lb} = 1, \ \forall l \in \mathcal{L} \tag{3.11a}$$

$$\sum_{l \in \mathcal{L}} y_{lb} \leq K_b, \ \forall b \in \mathcal{B} \tag{3.11b}$$

$$\sum_{b \in \mathcal{B}_l} c_{lb} y_{lb} \leq c_{\text{th}}, \ \forall l \in \mathcal{L} \tag{3.11c}$$

$$y_{lb} \in \{0, 1\}, \ \forall l \in \mathcal{L}, \ \forall b \in \mathcal{B}. \tag{3.11d}$$

Then, by encapsulating the problem of orchestrating the D2D links in a way that overall resource savings can be achieved to the above setting, we introduce the following bi-objective optimization solution:

$$\min \left\{ \sum_{b \in \mathcal{B}} \left(\sum_{l \in \mathcal{L}} y_{lb} \right)^2 \; ; \; \sum_{l \in \mathcal{L}} \sum_{b \in \mathcal{B}} \vartheta_{lb} y_{lb} \right\} \tag{3.12}$$

subject to:

$$\sum_{b \in \mathcal{B}_l} y_{lb} = 1, \; \forall l \in \mathcal{L} \tag{3.12a}$$

$$\sum_{l \in \mathcal{L}} y_{lb} \leq K_b, \; \forall b \in \mathcal{B} \tag{3.12b}$$

$$\sum_{b \in \mathcal{B}_l} c_{lb} y_{lb} \leq c_{\mathrm{th}}, \; \forall l \in \mathcal{L} \tag{3.12c}$$

$$y_{lb} \in \{0, 1\}, \; \forall l \in \mathcal{L}, \; \forall b \in \mathcal{B}. \tag{3.12d}$$

Without loss of generality, the two objectives are assumed to be equally important and an efficient balance between them is requested. Note that the first objective should be transformed in accordance with problem (3.8) to linearize the optimization problem.

The aim is to apply the weighted-sum (or scalarization [66]) method that combines two objectives into a normalized single-objective function (general form is $f_{tot} = f_1 + f_2$). For the formulation provided above (problem (3.12)), we denote by $f_1(\mathbf{s})$ the function that corresponds to the resource-aware balancing objective (first part) and by $f_2(\mathbf{s})$ the interference-aware part (second part) of the bi-objective problem. In order to make the unified objective tractable, the following normalizations must be applied:

$$f_{1_{norm}}(\mathbf{s}) := \frac{f_1(\mathbf{s})}{\max(K_b)} \in (0, 1], \tag{3.13}$$

$$f_{2_{norm}}(\mathbf{s}) := \frac{f_2(\mathbf{s})}{\sum \max(I^{\intercal})} \in (0, 1], \tag{3.14}$$

where $\max(K_b)$ in equation (3.13) is the maximum number of available RBs for all deployed BSs. Also, in (3.14), I is the matrix of size $|\mathcal{L}|\times|\mathcal{B}|$, where each element I_{lb} contains information about the number of the potential cellular interferers for the receiver of D2D link l in case it connects to a BS b; then, the denominator in equation (3.14) can be stepwise described as follows

$$I = \begin{pmatrix} \vartheta_{11} & \vartheta_{12} & \cdots & \vartheta_{1|\mathcal{B}|} \\ \vartheta_{21} & \vartheta_{22} & \cdots & \vartheta_{2|\mathcal{B}|} \\ \vdots & \vdots & \ddots & \vdots \\ \vartheta_{|\mathcal{L}|1} & \vartheta_{|\mathcal{L}|2} & \cdots & \vartheta_{|\mathcal{L}||\mathcal{B}|} \end{pmatrix}, \tag{3.15}$$

$$\max(I^{\intercal}) = \begin{pmatrix} \max\left(\vartheta_{11}, \vartheta_{12}, ..., \vartheta_{1|\mathcal{B}|}\right) \\ \max\left(\vartheta_{21}, \vartheta_{22}, ..., \vartheta_{2|\mathcal{B}|}\right) \\ \vdots \\ \max\left(\vartheta_{|\mathcal{L}|1}, \vartheta_{|\mathcal{L}|2}, ..., \vartheta_{|\mathcal{L}||\mathcal{B}|}\right) \end{pmatrix}^{\intercal}. \tag{3.16}$$

It equals to the summation of the maximum interferers for each D2D link when it associates with one of the candidate BSs.

Considering the above properties and by linearizing the first objective as shown previously, the problem is re-formulated to adapt to the weighted sum optimization technique. By applying the linearization methodology of (3.7) as shown before, the general form of the combinational objective function for the new optimization problem is as follows

$$\min \left[\; - w_1 \underbrace{\frac{z}{\max(K_b)}}_{f_{1_{norm}}} \; + \; w_2 \underbrace{\frac{\sum_{l \in \mathcal{L}} \sum_{b \in \mathcal{B}} \vartheta_{lb} y_{lb}}{\sum \max(I^{\intercal})}}_{f_{2_{norm}}} \; \right] \tag{3.17}$$

subject to:

$$z \leq \sum_{l \in \mathcal{L}} y_{lb}, \; \forall b \in \mathcal{B} \tag{3.17a}$$

$$\sum_{b \in \mathcal{B}_l} y_{lb} = 1, \; \forall l \in \mathcal{L} \tag{3.17b}$$

$$\sum_{l \in \mathcal{L}} y_{lb} \leq K_b, \; \forall b \in \mathcal{B} \tag{3.17c}$$

$$\sum_{b \in \mathcal{B}_l} c_{lb} y_{lb} \leq c_{\text{th}}, \; \forall l \in \mathcal{L} \tag{3.17d}$$

$$y_{lb} \in \{0, 1\}, \; \forall l \in \mathcal{L}, \; \forall b \in \mathcal{B} \tag{3.17e}$$

$$\sum_{i=1}^{N} w_i = 1. \tag{3.17f}$$

Constraint (3.17f) concerns the weights of the two ($N = 2$) objectives and enforces their summation to be equal to one, which is common practice in weighted-sum methodology [67].

3.4 D2D Resource Allocation

In this section, the second step of our two-stage approach is presented. Herein, an iterative randomized resource allocation (i-RRA) scheme for D2D communications based on the differentiated FFR is devised. Before introducing the RA algorithm, the following assumptions need to be considered. We make use of the fractional power control algorithm that sets the transmit power of a user device u associated with BS b according to [54]:

$$P_u = \min\{P_{\max}, \; 10 \log_{10}(M) + P_0 + \alpha PL_{ub}\}, \tag{3.18}$$

where $P_{\max}$ is the maximum transmit power of the device (24 dBm), M is the number of PRBs assigned to the device, P_0 is a normalized power value (in dB), α is the path loss compensation factor and PL_{ub} is the path loss between the transmitting UE u and its serving (associated) cell b. User index u corresponds to either a cellular UE or a D2D transmitter transmitting during the UL.

We recall from paragraph 3.2.2 that in a multi-cell scenario, due to the concurrent cellular and D2D transmissions, severe interference might deteriorate the rate performance of both user types. Cellular links can be harmed by multiple D2D active transmissions that utilize the same resource as well as by adjacent cells' cellular transmissions. On the other hand, D2D receivers suffer interference not only from other DUEs that transmit on the same channel but also from cellular transmissions of all cells. In order to calculate the achievable rate for both communication types, the received SINR at a D2D receiver (direct communication) and a BS (cellular communication) needs to be estimated:

$$\gamma_{ij} = \frac{P_i G_{ij}}{\sum_{q=1}^{|Q_j|} P_q G_{qj} + \sigma^2}. \tag{3.19}$$

In the case of cellular UL transmission, i corresponds to a transmitting CUE and j translates to the associated with user i BS. For a D2D pair, i is the transmitter and j the receiver. In the nominator, G_{ij} stands for the link gain in $i \to j$ transmission and P_i is the transmission power estimated according to eq. (3.18). In the denominator, the first factor represents the sum of the interference power from the other interfering signals. In detail, Q_j is the set of interfering nodes that utilize the same channel allocated for the $i \to j$ transmission , G_{qj} is the channel gain from interferer q to receiver j, and P_q is the transmission power of interferer $q \in Q_j$. Finally, σ^2 denotes the background/thermal noise power. Note also that, the mentioned link gains encapsulate slow channel fading (shadowing) impairments, with a shadowing standard deviation of 8 dB for both communication types.

According to the previous definitions, the received SINR for each transmission can be then mapped to achievable rate by using the Shannon capacity formula:

$$R_{ij} = B_{\mathrm{RB}} \log_2 (1 + \gamma_{ij}), \tag{3.20}$$

with B_{RB} being the RB bandwidth (180 KHz). Hence, the network's aggregate throughput is the summation of the achievable rates of all D2D and cellular communications.

3.4.1 Iterative Randomized RA algorithm (i-RRA)

Here, a low-complexity, iterative randomized algorithm is proposed which runs in a semi-centralized manner as follows: first, the cellular users of each cell are initially assigned with orthogonal RBs, depending on the area they are located in (inner or outer). Secondly, assuming that $\mathcal{N}$ RBs are available in an either an inner or an outer area, each BS will randomly allocate one RB per associated D2D pair according to the aforementioned FFR allocation logic and the area that the pair is located in. This RB will be then subtracted from the corresponding to each cell available RB pool. In case that all RBs are occupied (highly dense D2D scenarios), a cellular resource can be reused by more than one D2D link. This implementation will run for up to a designated number of iterations M. Then, the BSs cooperatively opt the best allocation pattern among all. The criterion for finally choosing the best resource allocation pattern to be applied is the total network throughput, estimated as the cumulative cellular and D2D transmissions' rates for all cells. The algorithmic steps are given in Algorithm 1.

The proposed algorithm's nature falls within the category of "embarrassing" parallel problems because iterations of the algorithm to explore the search space can be executed without requiring any communication between them [68]. Its computational complexity is $\mathcal{O}(M)$, which means that it only increases linearly

Algorithm 1: i-RRA ALGORITHM

Data: CUEs and DUEs' location coordinates,
$\mathcal{C}_b$: set of CUEs in cell b,
$\mathcal{L}_b$: set of D2D links associated with cell b (input from MOCA framework),
$\{N^b_{\text{Inner}}\}$: inner-region available RBs in each cell b,
$\{N^b_{\text{Outer}}\}$: outer-region available RBs in each cell b,
M: number of iterations.

for $b := 1$ *to* $|\mathcal{B}|$ **do**
- Allocate one unused (orthogonal) RB $\forall\ c \in \mathcal{C}_b$ from N^b_{Inner} or N^b_{Outer} depending on its location.

end
for $m := 1$ *to* M **do**
  for $b := 1$ *to* $|\mathcal{B}|$ **do**
  - $S_b \leftarrow \{N^b_{\text{Inner}}\} \cup \{N^b_{\text{Outer}}\}$.

  foreach $l \in \mathcal{L}_b$ **do**
  - Randomly allocate one RB r to $l \in \mathcal{L}_b$ from the corresponding S_b available RB pool.
  - Subtract assigned RB from the available pool: $S_b \leftarrow S_b \setminus \{r\}$.

  end
  end
- Compute $\forall l \in \mathcal{L}$ the achievable rate R_l and $\forall c \in \mathcal{C}$ the achievable rate R_c.
- $R_{total}(m) \leftarrow \sum_{b=1}^{|\mathcal{B}|} \Big(\sum_{c=1}^{|\mathcal{C}_b|} R_c + \sum_{l=1}^{|\mathcal{L}_b|} R_l \Big)$.

end
$T = \max\{R_{total}\} \rightarrow$ Maximum estimated Aggregate Throughput

Table 3.2: Simulation Parameters

Parameter	**Value**
User distribution	Uniform
Cell radius (d)	400 m
Number of macro cells	7
Number of CUEs per hexagonal cell	30
Total number of D2D links in hexagonal grid	150
Cellular path-loss model (PL_{CUE})	$128.1 + 37.6 \log_{10} d$
D2D path-loss model (PL_{D2D})	$148 + 40 \log_{10} d$
Max D2D link range (l_{n_1,n_2})	100 m
Maximum UE transmission power ($P_{\max}$)	20 dBm
Shadowing standard deviation	8 dB
Noise power spectral density	-174 dBm/Hz
System bandwidth (BW)	10 MHz

with the number of iterations M. To even reduce more the runtime, parallel processors can be used to distribute the computational complexity of running the algorithm for a big number of iterations.

3.5 Numerical Analysis

In this section, a set of numerical investigations is detailed in order to shed light on the performance of the *MOCA* framework to provide reliable D2D cell association. To realize this, D2D links and CUEs are uniformly distributed in a seven-hexagonal cell scenario to provide an indication of the performance in multi-cell network ecosystems. A key premise is that one RB will be allocated to each D2D and cellular link. The basic simulation parameters, similar to EU FP7 METIS project [69], are listed in Table 3.2. All results derive from Monte Carlo simulations in MATLAB.

For the performance evaluation of the proposed optimization framework a cost-based heuristic (CbH) cell association method for D2D UEs is devised. This technique greedily associates each D2D pair to the BS that averagely provides the best channel conditions to the two linked DUEs (Algorithm 2). Even though this method optimizes the coupling of the distributed D2D links according to path-loss based equation (3.3), it does not consider the BSs' limited resource availability which might lead to imbalanced cell association issues (thus, cases of over-loaded cells) in the long run. Basically, this algorithm runs in a centralized, sorted manner by sequentially associating each D2D link with the ideal BS to serve it. To avoid any over-utilization of the BSs, if all the available resources of a BS get occupied, the transmission of the respective DUEs is regulated from a competing BS that is less utilized.

Algorithm 2: COST-BASED HEURISTIC (*CbH*)

Data: DUEs' location coordinates, cost matrix $\mathbf{C} \in \mathbb{R}_{+^*}^{|\mathcal{L}|\text{x}|\mathcal{B}|}$, capacity vector $\mathbf{K} \in \mathbb{Z}_{+}^{1\text{x}|\mathcal{B}|}$.

$l = 1$;

while $l \leq |\mathcal{L}|$ **do**

- Find D2D link with *minimum* cost ($\min\{\mathbf{C}\} = c_{i,j}$, where i is the row (D2D link id) and j is the column (BS id)).

  if ($c_{i,j} \leq c_{th}$ & $K_j\ != 0$) **then**

  - Associate link i with BS j.
  - $K_j = K_j - 1$;
  - $c_{i,(1:|\mathcal{B}|)} = \text{Inf}$; → all i-th row values cannot be picked...

  else

  if ($c_{i,j} \leq c_{th}$ & $K_j\ == 0$) **then**

  - $c_{(1:|\mathcal{L}|),j} = \text{Inf}$; → all j-column elements cannot be picked...
  - Find *minimum* cost for i-th link and for $\{\mathbf{K} \setminus K_j\}$.
  - Update cost matrix ($\mathbf{C}$) and capacity vector ($\mathbf{K}$) accordingly.

  else

  - Link i cannot be associated (will associate in subsequent time epoch).

  end

  end

  $l = l + 1$;

end

Evaluation on MOCA-I

The overall picture though is very different and the gains are significant when a form of D2D control and resource utilization balancing is considered. For this simulation setting, we retain the same K_b values for each $b \in \mathcal{B}$ to be equal to the number of available D2D resources (i.e. 40) and assume that the optimization problem in (3.8) is solved for each instant and for the different number of D2D links. Starting from 70 D2D links in the hexagonal grid, for each different case another 20 is added and each BS associates with a number of D2D links, as restricted in (3.8). As mentioned before, each D2D link is then assumed to be allocated with one RB to satisfy its transmission needs. 1000 Monte Carlo simulations are executed to produce a statistical comparison of this problem's performance with the CbH method. Figure 3.5 presents the normalized minimum RB availability that is achieved by using these two methods in the cases of hexagonal and PPP-Voronoi based cell layouts. By normalized minimum RB availability we mean the percentage of unallocated RBs among the deployed BSs that basically translates to the most utilized BS. On average, and for the case of hexagonal multi-cell environments, 12% more resources are available when using MOCA-I. In the same figure, a statistical maximum 19% of minimum RB availability is shown for the case of 90 D2D links. Compared to it, in random deployment scenarios where the BSs' locations follow PPP modeling, the estimated RB availability is distinctly higher. This can be explained by the randomness of the cell shapes that, in some cases, allows for having multiple cell-edge links in the borders of more than two cells, and consequently, more BSs are candidates for association with them. According to the Monte Carlo based statistics, MOCA-I achieves an almost 10% increase in terms of resource availability, compared to CbH method, and a peak performance gain of 15% in the case of 130 links.

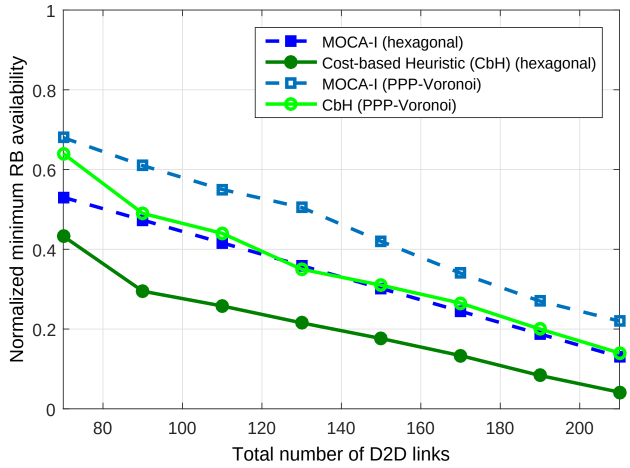


Figure 3.5: Normalized minimum RB availability for MOCA-I and CbH techniques in relation to the number of D2D links.

Evaluation on MOCA-II

For this case study, we focus on high-congestion network episodes, where some resources have to be inevitably reused for more than one D2D pair within the network. When congestion level is mentioned we refer to the number of D2D transmission requests in proportion to the BSs' available resources. Without loss of generality, we assume that each available RB per BS is assigned with an integer value that is randomly picked (from $[0, 2]$) and indicates the number of times this RB is already used. Regarding the bi-objective optimization setting (problem (3.10)) described in subsection 3.3.2, the first objective function is the minimization of the path-loss based cost that we proved is only slightly better compared to the CbH method. Hence, we solve this problem for the case of RB reuse avoidance optimization (second objective function) and for highly dense scenarios. To this end, Figure 3.6 proves that for different traffic cases, this solver achieves better performance compared to an average utilization agnostic method that assigns resources randomly picked from the available pool of RBs. To be more specific considering the latter, it does not take into account the existing RB reuse cases; in contrast, it randomly gets allocated with a random resource depending on the BS it is associated with. On the other hand, we also consider

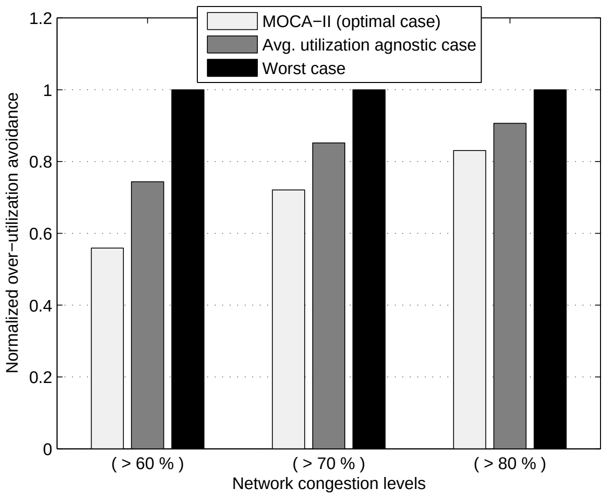


Figure 3.6: Normalized RB over-utilization avoidance optimization compared to average utilization agnostic and worst case scenarios.

the worst case where the under study D2D links would be assigned with the over-utilized RBs. As depicted, for example in the case of hexagonal deployment, the proposed method succeeds in utilizing less used RBs and outperforms the worst case as well as the average utilization agnostic case in a percentage of 45% and almost 28% over-utilization avoidance gain, respectively. Specifically, this performance gain can be depicted in the same figure for the case of more than 60% of overall network congestion. We note that similar behaviour is observed for more congested instances (i.e. $> 70\%$ and $> 80\%$) where the achieved gain remains significant. It has to be also noted that, because the number of distributed D2D links as well as the input values ρ (already assigned RBs) are the same for both hexagonal and PPP-Voronoi tessellation layouts, the differences in terms of RB over-utilization are negligible between each other.

Evaluation on MOCA-III

Problem (3.17) aims at efficiently performing resource-aware balancing while taking into account the existence of potential cellular interferers to the D2D communications. In that sense, the interference-aware part adds a useful decision-making dimension to the optimization setting and mainly relates to the statistical chance of a link to interfere with a CUE. In essence, it chooses the association of a D2D link with the BS that will result in the lowest probability of interfering with a closely located CUE. Figure3.4 represents the interference region of a CUE (orange-dashed circular area in the figure) as the area under which cellular users might interfere with a DUE receiver. In specific, this area can be specified by the transmitting CUE's location (centre of the circle) and a radius that is defined as the range of the interference. For approximation, the interference range for each CUE might vary and can be approximated by the multiplication of the respective cellular link range with a random variable $\beta \geq 1$ [70]. Without loss of generality, we consider an upper limit for this variable and assume that $\beta \in [1, 2]$. By considering uniform distribution for both CUEs and DUEs, it might be possible to have an equal number of potential cellular interferers when a D2D link associates with two different BSs. A representative toy example is shown in Figure 3.7: we consider a cell-edge D2D link that its transmitter is located on the edge of a cell's inner-region and the receiver being in a designated range from it (we considered 100 m). Then, we shift the link on a straight line with a step of 50 meters towards the neighbouring competing BS to look on the number of possibly interfering nodes if the link is associated to each one of the two BSs. As expected, when the link crosses the two-cell limit and is positioned almost in the middle of the distance between the two BSs, the number of the interferers is equal (in our case, this happened in a 150-meter shift from the initial position). In such scenarios, it is the balancing objective that would guide the solution.

Finally, we solve the problem (3.17) in order to obtain an efficient trade-off of the two objective functions. It has to be observed that these two objectives are non-conflicting and therefore we are not considering optimal trade-off operating points on a Pareto front[3]. The two objective functions are normalized according

[3]Pareto front is the set of Pareto optimal outcomes in multi-objective optimization problems (source: https://en.wikipedia.org/wiki/Multi-objective˙optimization)

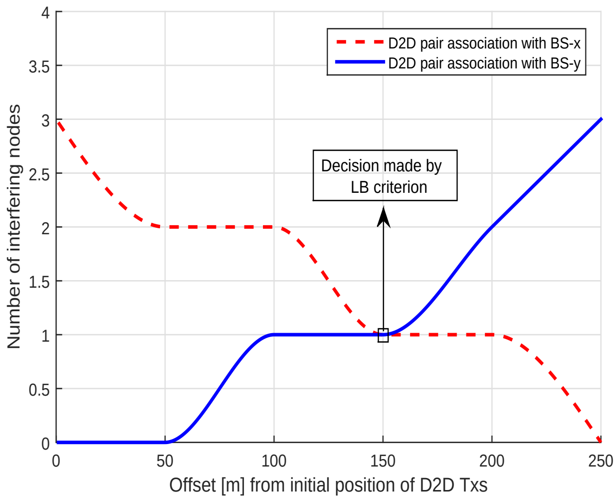


Figure 3.7: Example of potential CUE interferers for a D2D link. The depicted intersection point means equal interference probability for each possible association. The stated LB (load balancing) criterion means that, because the probability is equal in this case, the association balancing function (f_1) will decide for the association of this pair to a BS by contributing on providing as much as possible a balanced orchestration of the distributed links.

to the analysis detailed in subsection 3.3.3. The obtained result, as shown in Figure 3.8, indicates that there is linear non-conflicting relation of the two objectives. Therefore, as we can clearly observe, a decrease to the second objective function can be interpreted as an increase to the max-min output of the balancing factor (and vice versa) in order to retain the solution of the optimization problem in its optimal value. In detail, it is the nominal practice to chose weights that their summation equals to one. Therefore, by applying a stepwise increase of the w_1 weighting factor and a corresponding decrease to w_2, we show how the two objectives behave. Since the objectives are non-conflicting and their relative importance can be deemed as equal in general, the decision maker (i.e. network operator) can tune the aforementioned weights considering the relative magnitude of the objectives in order to achieve different network operating points.

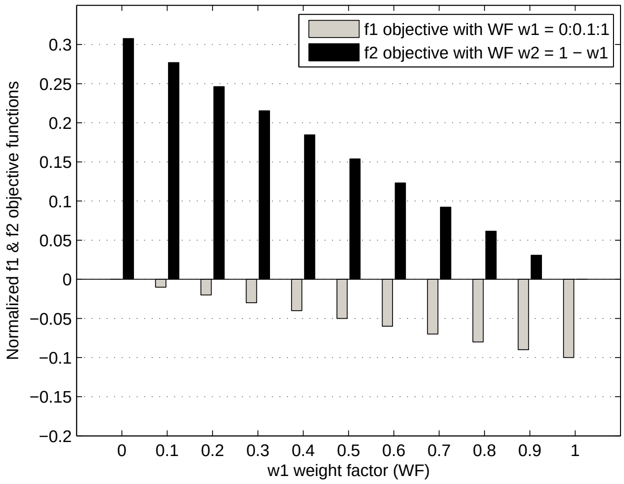


Figure 3.8: Weighted sum method for MOCA-III bi-objective problem (sum constraint $w_1 + w_2 = 1$ is being applied).

Throughput performance

In our proposal, the solution of a CAS optimization problem will work as feed for the resource allocation technique described in 3.4.1. Among the proposed optimization formulations, we opt to use MOCA-I throughout the rest of the simulations as it provides the lowest running time complexity out of the three MOCA proposals. Indicatively, solving the three problems with the same CPU (INTEL(R) CORE(TM) i7-6500 @ 2.50 GHZ / 8 GB RAM), MOCA-I runs in 0.8 seconds, whereas MOCA-II and MOCA-III run in 1.4 and 2.3 seconds, respectively. By running this resource balancing-oriented setting, each D2D pair in the topology will be thereafter associated with a BS, aiming at contributing to the maximization of the minimum RB availability of the network.

Initially, we will compare the performance of the joint optimization problem presented in 3.2.4 with a set of different allocation techniques that use MOCA-I to decide over the cell association pattern for D2Ds. First, our *i-RRA* proposal that is detailed in 3.4.1. Second, a differentiated FFR algorithm (for simplicity we abbreviate it as *dFFR*) [64], which uses the same FFR-based allocation rationale for D2D links but requires the cell outer-region DUEs to use the resources that cannot be used by the inner-region DUEs in order to guarantee the latter's welfare, providing thus a form of prioritization. Third is the soft FFR-based algorithm (*sFFR*) [65]. In that case, D2D and cellular UEs, being located in the same region (inner or outer), are able to utilize resources from the same available RB pool (e.g. in the center cell's inner region, *F1* is the available RB pool, whereas for outer region is *F2*). This however can be potentially harmful for dense scenarios, because the existence of multiple D2D links will potentially lead to over-reuse of some of the limited resources (especially for cell-edge users) and, thus, performance might get degraded. Lastly, we apply a baseline random allocation algorithm, where D2D UEs can be assigned with a random RB from the whole frequency band that implies a rather unexpected performance. Fig. 3.9 shows the D2D sum-rate performance of all pre-mentioned resource allocation schemes which is upper bounded by the joint optimization problem's solution. As already discussed, the latter is solvable in the case of a small total number of D2D links that are randomly distributed in space because resource overlaps can be then avoided (no RB reuse by multiple DUEs). We use again the weighted sum method presented in III.C to run the optimization problem by assigning the weight factors (WF) w_1 and w_2 with 0.5. It is shown that the joint optimization solution outweighs the above algorithms by almost 17%, 63%, 132% and 104% on average in terms of sum-rate, respectively, for the case of hexagonal grid scenario. As expected, with the increase of the number of D2D links, the sum-rate improves proportionally in most of the depicted cases. However, in the last one (40 D2D links), only the joint optimization retains this increasing tendency by maximizing the total achievable throughput for DUEs, whereas in the rest of the algorithms, the developed interference due to the increasing number of users leads to a slight performance degradation. Also, even though the achieved D2D sum-rate performance drops almost in half in PPP-Voronoi based deployments, the proposed method lays in between the joint optimal (but high complexity)

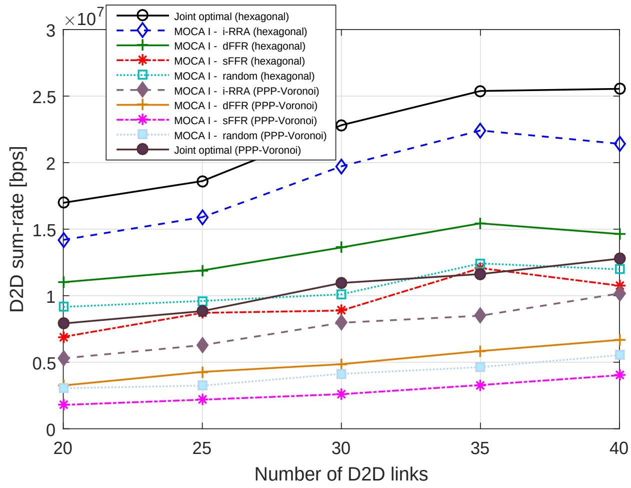


Figure 3.9: D2D sum-rate comparison. In the joint optimal case a weight factor of 0.5 ($w_1 = w_2 = 0.5$) is used for both objective functions in (3.6).

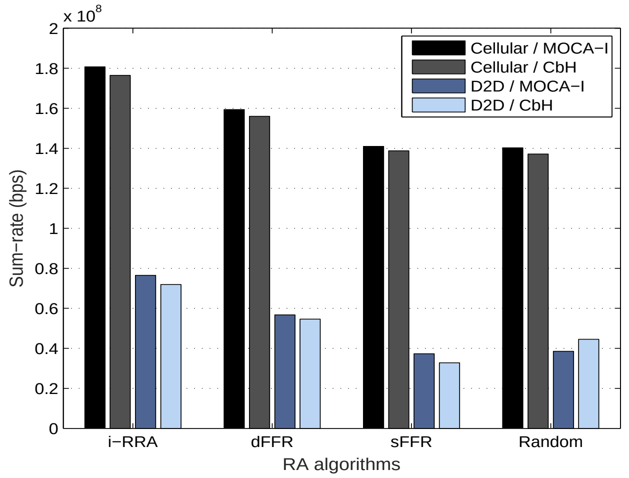


Figure 3.10: Throughput performance for cellular and D2D connections in hexagonal based cellular networks.

method and the compared baseline techniques.

For the rest of the performance evaluation, we investigate the resource allocation problem from a high D2D-related density point of view (150 D2D pairs, uniformly distributed). Figure 5.5 highlights the maximum sum-rate performance of the four resource allocation techniques for both cellular and direct transmissions. For this case study, we also compare the presented CAS methods (MOCA *vs* CbH) in order to visualize any effect on the throughput performance other than resource-aware utilization for the network. The difference per case is low,

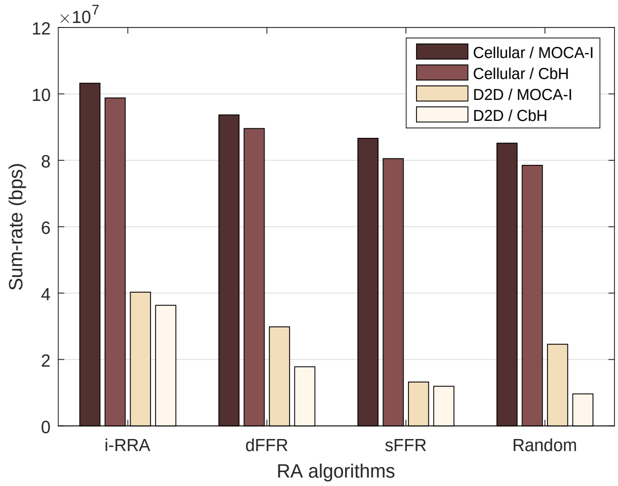


Figure 3.11: Throughput performance for cellular and D2D connections in PPP-Voronoi deployments.

however if we consider the resource utilization savings already presented in this section, on top of 12% minimum RB availability, the i-RRA algorithm based on MOCA-I CAS gives a slight improvement of almost 2.5% compared to the same algorithm with CbH. The aggregate throughput gain is more visible for D2D communications where 6.5% improvement is achieved through the MOCA-I / i-RRA two-stage implementation. By leveraging the MOCA-I CAS technique as the first step of the solution, the i-RRA for D2D communications outperforms the dFFR, sFFR and random RA techniques in a percentage of 35%, 105% and 98% respectively. In addition, the cellular performance follows the same trend, as the i-RRA algorithm is better than the rest of the methods for a percentage of 13%, 28% and 29%, respectively. Complementary to it, Fig. 3.11 showcases the adaptability of the proposed methodology to more realistic deployments, represented by Voronoi cell tessellation. Again, a clear sum-rate performance improvement can be observed for both cellular and D2D users when leveraging the MOCA-I association method over the CbH one. Also, another representative fact is the further increased D2D sum-rate gain when using the i-RRA algorithm, while it outperforms the rest of the methods in almost 36%, 204% and 65%, respectively.

Finally, Figure 3.12 shows the CDF of the achievable rates for all users. Like

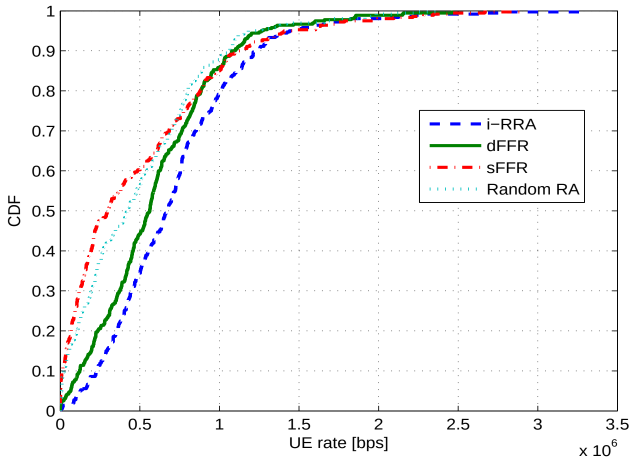


Figure 3.12: UE rate CDF in hexagonal layouts based on MOCA-I.

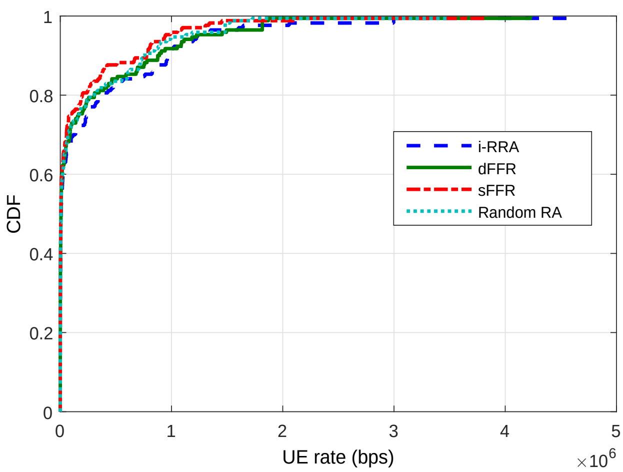
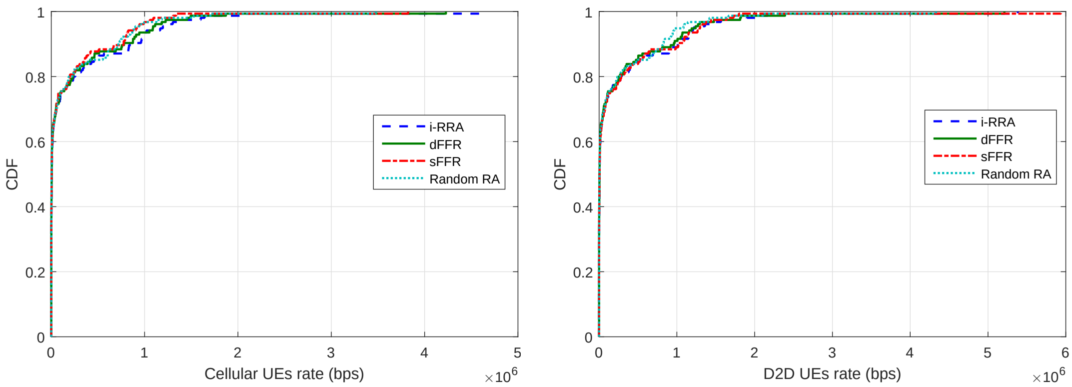


Figure 3.13: UE rate CDF in PPP-Voronoi deployments based on MOCA-I.



(a) Cellular UEs rate CDF **(b)** D2D UEs rate CDF

Figure 3.14: Drilled-down view of cellular and D2D UEs rate CDF in PPP-Voronoi deployments based on MOCA-I.

before, MOCA-I is considered to be the decision mechanism used for the D2D-BS association. The i-RRA algorithm proves its supremacy and entails better throughput performance. In the 50$^{\text{th}}$ percentile, the i-RRA algorithm achieves a 17%, 110% and 48% better performance compared to dFFR, sFFR and Random methods, respectively. Considering the 90$^{\text{th}}$ percentiles, over 27% better rate performance of the i-RRA over the other RA techniques is shown. Again, same behavior is observed when investigating the UEs' throughput performance from the Voronoi tessellation point of view (Fig. 3.13), even though the average user throughput drops as compared to the hexagonal case. The latter can be further drilled-down as in Fig. 3.14 to separately indicate the performance of cellular and D2D UEs rate in terms of CDF depiction, where it is shown that, yet again, i-RRA outperforms the rest of the techniques in a clear fashion.

3.6 Summary

In this section, a two-stage approach for achieving spectrum efficiency and enhanced throughput for D2D enabled networks is proposed. First, a set of ILP optimization problems (namely MOCA) is devised, aiming to optimize the cell association (CAS) aspect for D2D communications with respect to resource limitations, interference and network congestion episodes. The proposed set of optimization problems is amenable for a centralized implementation, something that could potentially be in-line with emerging cloudified RAN-based mobile networks. Then, based on the output of the cell association solution, a low-complexity iterative randomized algorithm for D2D communications that considers different RB pools for CUEs and DUEs is applied. The proposed framework is compared with baseline methods as well as related works in the literature. In terms of resource efficiency, the CAS optimization entails a balanced association of the distributed D2D UEs to the deployed BSs that can be interpreted as valuable resource savings and network capacity ease; over 12% of resource savings can be admitted by this method compared to a path-loss based heuristic one. Then, the proposed iterative randomized algorithm, called i-RRA, provides a fast and effective solution in terms of sum-rate performance when compared to other existing algorithms. Over 34% of D2D sum-rate improvement can be realized via i-RRA, with a non-degrading cellular achievable performance.

Chapter 4

Bio-Inspired Resource Management for Relay-Aided D2D Communications

4.1 Introduction

D2D communication emerges as an attractive way to tackle the dramatic increase in traffic and shortage of spectrum in cellular networks by capitalising on the proximity of UEs to each other. D2D as an underlay in cellular networks enables the reuse of the spectrum assigned for cellular communications. It also allows the offloading of cellular traffic and enables more reliable and high throughput links between users in close proximity. For this reason, and following the prediction for further densification in future 5G networks, D2D is expected to play a principal role in spectrum and resource management since in several cases the number of D2D connections can be very high and the resources would need to be carefully controlled. However, in a D2D-enabled network some challenges need to be addressed in order to get the full benefit of this technology. Firstly, the potential D2D UEs may not be in close proximity which may render the establishment of a reliable connection between the D2D UEs challenging. In addition, the high spectral efficiency of underlay operation comes at the price of high levels of interference to and from CUEs which could jeopardize the QoS of D2D as well as cellular UEs.

In this chapter, we study the joint resource allocation for cellular and relay-aided underlay D2D communications. We consider that cellular/D2D UEs could act as relay nodes in order to enhance the link quality between D2D UEs that are far from each other or the channel quality between them is poor. Additionally, we consider that all UEs are eligible to either communicate directly with their peer or via relaying. Our proposal considers the use of bio-inspired genetic algorithms (GAs) [71] in order to find a near-optimal allocation of resources for cellular and D2D UEs that can potentially achieve the maximum sum-rate within the network. GA has proven to be an efficient approach for limiting the solution search space via its probabilistic methodology (as analyzed in the next paragraphs) and achieving near-optimal solutions to optimization problems.

4.1.1 Closely related work

Bio-inspired genetic algorithms (GAs) [71] have become a popular approach in solving resource allocation problems in wireless networks [72, 73, 74] mainly because of their versatility, scalability and computational simplicity which make GA a very attractive method to solve the resource allocation problem as it will be detailed below. Resource allocation for D2D communications has been extensively studied within the literature. In [75], a proportionally fair utility maximization approach is used to allocate resources to both DUEs and CUEs. In [76] the mode selection and resource allocation problems for underlay D2D communication are investigated and solved using particle swarm optimization. Further, an efficient graph-theoretical approach is proposed in [77] to perform channel allocation for DUEs. Resource allocation in relay-aided D2D scenario has been studied in [47].

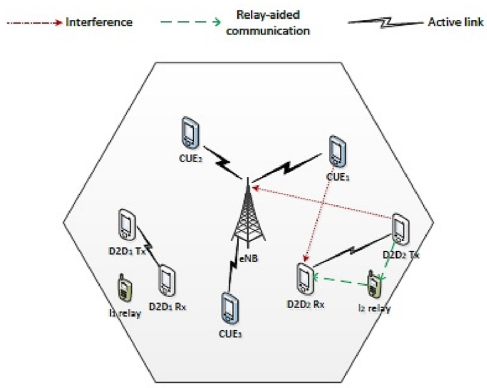


Figure 4.1: Uplink scenario of relay-aided D2D communications as an underlay to the cellular network.

However, [47] considered that all traffic flows are routed through L3 standard relays whereas in our study the choice of direct or relayed D2D communication based on the achievable rate is part of the optimization problem.

4.2 Problem Definition

The resource allocation problem in cellular networks is a widely studied area that falls within the nature of NP-hard problems which cannot be solved in real-time. Popular integer relaxation methods have been applied to reduce its time complexity but do not render it a real-time solution for network operators. In this section, we define important preliminary notations and parameters that will help us further formulate the relay-aided D2D/cellular resource allocation optimization setting and pave the way for our proposal.

4.2.1 Preliminaries

First, we consider the UL case scenario of D2D underlaying a cellular network where interference patterns from a CUE to a receiving DUE and from the transmitting DUEs to the CUE UL transmission, as depicted in Fig. 4.1. In this figure, interference exerted from the cellular user CUE_1 towards the $D2D_2$ pair and vice versa might be destructive not only for the reliability of the link, but also for the aggregate network throughput. Therefore, these two transmissions should occupy different RBs to avoid mutual interference. Additionally, we assume that cellular users are directly transmitting to the serving BS, whereas the communication mode between two DUEs can be either direct or with the help of a closely located relay that lays within the geographical serving area of a macro cell. An important assumption is that only one proximate UE to a D2D pair can be used as its relay to help the transmission. Now, before we detail the problem formulation, we need to define the following sets:

- $\mathcal{K} = \{1, 2, \ldots, K\}$: set of available RBs.
- $\mathcal{L} = \{1, 2, \ldots, L\}$: set of D2D links.
- $\mathcal{C} = \{1, 2, \ldots, C\}$: set of cellular links.
- $\mathcal{M} = \{1, 2, \ldots, M\}$: set of relays.

Also, in order to formulate this problem, we need to further define the decision variables of the optimization setting that will be valued according to the assignment (or not) of a RB to a specific user, either for a cellular, a direct or relayed D2D communication. The binary variables that correspond to CUEs, relayed D2D or direct D2D RB allocation are defined by (5.3)-(4.3) respectively.

$$x_c^k = \begin{cases} 1, & \text{if CUE } c \in \mathcal{C} \text{ transmits on RB } k \in \mathcal{K} \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

$$y_{ij}^k = \begin{cases} 1, & \text{if DUE (relay) } i \text{ sends to relay (DUE) } j \text{ via } k \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

$$z_l^k = \begin{cases} 1, & \text{if D2D pair } l \text{ communicates directly with RB } k \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

We consider a deterministic model where the SINR between nodes i and j over RB k, denoted as γ_{ij}^k, can be expressed as

$$\gamma_{ij}^k = \frac{P_i G_{ij}}{I_{j,k} + \sigma^2}, \tag{4.4}$$

where $I_{j,k}$ is the interference received by user j over resource block k, P_i is the transmitted power of node i, G_{ij} is the link gain between node i and j, and lastly, σ^2 is the power of background/thermal noise. The D2D interference to the UL transmission of CUE c to BS b over RB k is denoted by $I_{c_b,k}$ and is given by

$$I_{c_b,k} = \sum_{l \in \mathcal{L}} \left(P_l G_{lb} z_l^k + \sum_{m \in \mathcal{M}} (P_l G_{lb} + P_m G_{mb}) y_{lm}^k \right). \tag{4.5}$$

We note that, for the rest of the paper, the i, j indexes in G_{ij} (or y_{ij}) correspond to the transmitter and the receiver respectively. Also, from now on, we use y_{lm} to refer to the link between a DUE of pair l to relay m and vice versa. Depending on the position, if the notation l is the first sub-index, it implies the transmitter of this D2D pair, whereas in the other case it refers to the receiver of it. Also, it is clear that $y_{lm}^k = y_{ml}^k$ for a D2D pair l and a relay m.

The uplink channel rate of the cellular user c over resource block k, denoted by R_c^k, is given by

$$R_c^n = B_{RB} \log_2 \left(1 + \frac{P_c G_{cb} x_c^k}{I_{c_b,k} + \sigma^2} \right), \tag{4.6}$$

where B_{RB} is the RB bandwidth (180 kHz), P_c is the transmit power of CUE c. Finally, the overall data rate for this CUE is

$$R_c = \sum_{k \in \mathcal{K}} R_c^k. \tag{4.7}$$

Similarly, the interference affecting the D2D receiver of a pair l can be from the cellular user c or the other DUEs/relays that are transmitting over the same resource. If the resource block k is assigned to l, the received interference power for l, denoted by $I_{l,k}$, is given by

$$I_{l,k} = \sum_{c \in \mathcal{C}} P_c G_{cl} x_c^k \; + \sum_{i \in \mathcal{L} \setminus \{l\}} \left(P_i G_{il} z_i^k + \sum_{m_i \in \mathcal{M}} (P_i G_{il} + P_{m_i} G_{m_i l}) y_{im_i}^k \right). \tag{4.8}$$

The rate of the direct D2D communication of link l is then given by

$$R_{\text{direct},l} = B_{RB} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{P_l G_l z_l^k}{I_{l,k} + \sigma^2} \right), \tag{4.9}$$

where G_l is the channel gain for the D2D pair l transmission.

If the relay-based communication is used for D2D pair l via a relay m, the link capacity of the first and second hop respectively are given by

$$R_{lm}^k = B_{RB} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{P_l G_{lm} y_{lm}^k}{I_{m,k} + \sigma^2} \right), \tag{4.10}$$

$$R_{ml}^k = B_{RB} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{P_m G_{ml} y_{ml}^k}{I_{l,k} + \sigma^2} \right), \tag{4.11}$$

where $I_{m,k}$ is defined as the interference power from CUE and the other D2D users exerted to relay node m, and is given by exchanging the subscript l with m in (4.8).

Lastly, if we consider that relays are operating in full-duplex (FD) mode (relay nodes are able to receive and transmit data in same frequency bands and time slots) in amplify-and-forward communication [78], R_m^k, which is given below, denotes the total achieved rate for a relay-aided D2D communication over RB k where m refers to the relay that assists the considered D2D pair l.

$$R_m^k = \min\{R_{lm}^k,\ R_{ml}^k\}. \tag{4.12}$$

4.2.2 Problem formulation

We define the sum-rate maximization problem in an LTE scenario where D2D UEs underlay cellular communications:

$$\max \sum_{k\in\mathcal{K}} \left[\sum_{c\in\mathcal{C}} R_c^k x_c^k + \sum_{l\in\mathcal{L}} \left(R_{\text{direct},l}^k z_l^k + \sum_{m\in\mathcal{M}} R_l^k y_{lm}^k \right) \right] \tag{4.13}$$

subject to:

$$\sum_{k\in\mathcal{K}} R_c^k x_c^k \geq R_{th},\ \forall c \in \mathcal{C} \tag{4.13a}$$

$$\sum_{k\in\mathcal{K}} \left(R_{\text{direct},l}^k z_l^k + \sum_{m\in\mathcal{M}} R_m^k y_{lm}^k \right) \geq R_{th},\ \forall l \in \mathcal{L} \tag{4.13b}$$

$$\sum_{c\in\mathcal{C}} x_c^k = 1,\ \forall k \in \mathcal{K} \tag{4.13c}$$

$$\sum_{k\in\mathcal{K}} x_c^k = 1,\ \forall c \in \mathcal{C} \tag{4.13d}$$

$$\sum_{k\in\mathcal{L}} \left(z_l^k + \sum_{m\in\mathcal{M}} y_{lm}^k \right) = 1,\ \forall l \in \mathcal{L} \tag{4.13e}$$

$$x_c^k, y_{lm}^k, z_l^k \in \{0,1\},\ \forall k \in \mathcal{K},\ l \in \mathcal{L},\ m \in \mathcal{M},\ c \in \mathcal{C}. \tag{4.13f}$$

Constraints (4.13a), (4.13b) restrict the rate to be above a predefined threshold for all communications, i.e. direct, relayed D2D and cellular connections. Following the milestones of LTE, (4.13c) imposes the orthogonal assignment of the cellular users. Also, constraint (4.13d) signifies the allocation of each cellular user c with a single RB, whereas (4.13e) applies the same RB limitation for the D2D communication and also implies that only one relay can be potentially assisting each D2D link. Thus, the role of the binding variables z, y in the latter constraint is to restrict each D2D to communicate only in direct or relay mode and can be considered as a logical OR set of constraints. Set aside the NP-hardness of the above optimization problem, and assuming that, in the best case scenario, cellular users are allocated in a static way with specified RBs, solely the objective function's complexity is lower bounded by $\mathcal{O}(L!)$ in the case that L links are able to satisfy their transmission needs with one RB, where $L \leq K$. This insinuates that a potentially huge growth in various networking topologies, where either the number of users or the number of available radio resources (higher system bandwidth) is enhanced, translates to increased complexity. To this end, the need for introducing lower complexity solutions is immense.

4.3 Genetic Algorithm

GA is one of the most popular bio-inspired algorithms and is used to tackle real world NP-hard optimization problems. In general, bio-inspired algorithms imitate the natural evolution of biological organisms to provide a robust, near optimal solution for various problems. GA is inherently an evolutionary process that involves chromosome encoding, population initialization, fitness function depiction, crossover and selection mechanisms. These operations will be briefly explained in Section 4.3.1. A detailed analysis of GAs can be found in [71]. Initially, we introduce the following two important definitions.

Problem mapping: The first step in solving the resource allocation problem using GA is to establish a mapping between them. Since our problem space corresponds to CUE or DUE channel allocation, an integer based chromosome coding mechanism will be used. Based on this, each individual can directly map to a potential channel allocation for CUEs and DUEs where a channel allocation for a UE is represented by a chromosome; a set of chromosomes forms an individual. The initial population consists of a certain number of individuals, denoted by V. A common method to initialize the population is to randomly generate the chromosomes of each individual. In addition, the feasibility of each individual should be ensured to accelerate the convergence process. Thus, we first randomly generate two feasible vectors for each node, according to the representation scheme. Once all vectors are available, they will be combined to form a feasible individual with length equal to $(C + L + M)$. This is repeated until V individuals are generated. The formed population then acts as the very first generation that kicks off the subsequent evolving steps.

$$
\begin{aligned}
f = \sum_{k\in\mathcal{K}} \left[\sum_{c\in\mathcal{C}} R_c^k x_c^k + \sum_{l\in\mathcal{L}} \left(R_{\text{direct},l}^k z_l^k + \sum_{m\in\mathcal{M}} R_m^k y_{lm}^k \right) \right] \\
+ \sum_{c\in\mathcal{C}} \alpha_1 \min \left(R_{th} - \sum_{k\in\mathcal{K}} R_c^k x_c^k, 0 \right) \\
+ \sum_{l\in\mathcal{L}} \alpha_2 \min \left(R_{th} - \sum_{k\in\mathcal{K}} \left(R_{\text{direct},l}^k z_l^k + \sum_{m\in\mathcal{M}} R_m^k y_{lm}^k \right), 0 \right)
\end{aligned}
\tag{4.14}
$$

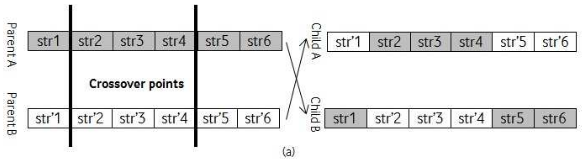
Fitness function: To this end, we firstly need to interpret the objective of the optimization problem in (4.13) to a fitness function that evaluates the quality of a given individual. In this case, to formulate this we apply a penalty function to ensure that constraints (4.13a) - (4.13b) are satisfied. In addition, the D2D mode selection (i.e. direct or relayed) is also optimized during fitness evaluation. The fitness function is defined in (4.14).

4.3.1 GA operation

1) Selection: An operation used for choosing individuals to participate in reproduction. In this study, the roulette wheel selection model is used where the chosen probability is proportional to the individual fitness evaluation function. Its selection probability for individual i is defined as

$$p_i = \frac{f(i)}{\sum_{i\in V} f(i)}. \tag{4.15}$$

2) Crossover and mutation: Crossover mixes the current solution so as to find better ones whereas mutation helps the GA avoid local optima. In the context of this chapter and for the simulation results, one and two points (OP



C_1	C_2	D_1	D_2	D_3	L_1	L_2	L_3
2	3	1	3	6	4	3	3

C_1	C_2	D_1	D_2	D_3	L_1	L_2	L_3
2	3	1	4	6	2	3	3

Figure 4.2: (a) Two-point crossover example. (b) Mutation example.

and TP) crossover cases are considered. An example of a two-point crossover is illustrated in Fig. 4.2(a). The mutation operation works by randomly making minor changes in the chromosomes after the crossover operation is performed. In our algorithm, we view each chromosome as a single gene. We define a trivial probability p_v as the likelihood of a gene to mutate. If a gene is determined to mutate, one digit of the vector will be randomly selected and replaced with a different value as shown in Fig. 4.2(b).

3) Replacement: After generating a new population, an elitist-based replacement model is employed to modify the old population with a certain number of new individuals. The worst individuals in the parental population are replaced by their children in the next generation.

The algorithm works as follows: an initial population is initialized. Then, the reproduction process starts, including mutation and crossover. The worst individuals are replaced with fitter ones based on the fitness function and this process is repeated until the maximum number of generations is reached. Considering the run-time performance of the GA, it is dependant on the three mentioned procedures. It is proven that GA scales well in terms of time complexity compared to ILP problems that are unable to run for highly dense topologies [79]. Briefly, the complexity of GA algorithms estimated as $\mathcal{O}(V * G * \mathcal{O}(\text{Fitness Function}) * ((P_{cr} * \mathcal{O}(\text{Crossover})) + (P_{mut} * \mathcal{O}(\text{Mutation}))))$, where V (population size), G (number of generations), P_{cr} (crossover probability) and P_{mut}

(mutation probability) are constants which further simplifies the complexity to $\mathcal{O}(\mathcal{O}(\text{Fitness Function}) * (\mathcal{O}(\text{Crossover})) + \mathcal{O}(\text{Mutation})))$.

4.4 Heuristic Algorithm

Herein, an algorithm that focuses on providing an enhanced rate performance for D2D users with respect to cellular throughput is devised. A basic assumption is that cellular users are initially allocated with orthogonal resources to satisfy their UL transmissions. Further, we iterate over all D2D links and pre-calculate for each one of them their potential rate performance (according to Shannon capacity formula) on each RB, based on the interference from cellular UEs. Then, we identify the best combination of D2D UE and RB that gives the maximum among all rate as a starting point. Recall that the maximum rate of a UE on a specific RB can result from either direct or relayed communication. Then, we update the rate matrices ($\mathbf{l}_{\mathrm{v}}$ for direct and $\mathbf{m}_{\mathrm{v}}$ for relayed transmission) with the former step's allocation and iterate over all UEs by taking into account the interference deriving from this RB assignment. Last, after all UEs are served, we estimate the rate that each UE achieves through the final allocation pattern and consequently the overall throughput. The algorithmic steps are analytically shown in Algorithm 3.

4.5 Performance Evaluation

In this section, a set of numerical investigations is presented to evaluate the performance of the GA-based resource allocation method. The results derive from Monte Carlo simulations of 100 iterations, implemented in Matlab. It has to be noted that the path-loss considered for a relay-enabled communication is the same as in the rest of the communication modes as each relay node is a D2D-enabled UE. Also, one RB is assumed to be assigned for each transmission. The

Algorithm 3: Sum-rate maximization algorithm

Input : $\mathcal{C}$, $\mathcal{L}$, $\mathcal{M}$, $\mathcal{K}$ (with their corresponding cardinalities C, L, M, K) / users' location.
Output: Aggregate throughput: R_{tot}

```
for c := 1 to C do
    - allocate random orthogonal RB k to user c; K_cellular = K_cellular − {k};
end
for i := 1 to L do
    for k := 1 to K do
        - calculate l_v(i, k);
        - calculate m_v(i, k);
    end
    l_v^max(i) = max(l_v(i, :));
    m_v^max(i) = max(m_v(i, :));
end

S = zeros(L, 2);
j = 1;
while j ≤ L do
    find < l, k > combination that gives the maximum rate among all
      elements in l_v^max and m_v^max matrices;
    S(l, :) = [l, k];
    repeat
        - update the rates on the assigned RB k ∀u ∈ L − {l} for both l_v,
          m_v;
        - update l_v^max(u) & m_v^max(u);
        l_v(l, :) = 0; m_v(l, :) = 0;
    until all matrices' rows are updated
    j = j + 1;
end
for l := 1 to L do
    - calculate achieved rate for direct or relayed D2D comm. for user
      l (R_l);
end
for c := 1 to C do
    - calculate achieved rate for cellular link c (R_c);
end
R_tot = Σ_{c∈C} R_c + Σ_{l∈L} R_l;
```

Where: $\mathcal{K}_{cellular} = \mathcal{K}_{cellular} - \{k\}$; $\mathbf{l}_v(i,k)$, $\mathbf{m}_v(i,k)$, $\mathbf{l}_v^{max}(i) = \max(\mathbf{l}_v(i,:))$, $\mathbf{m}_v^{max}(i) = \max(\mathbf{m}_v(i,:))$, $\mathbf{S} = zeros(L,2)$, $\forall u \in \mathcal{L} - \{l\}$, $R_{tot} = \sum_{c\in\mathcal{C}} R_c + \sum_{l\in\mathcal{L}} R_l$.

Table 4.1: Simulation Parameters

Parameter	**Value**
User distribution	Uniform
Macro cell radius	250 m
D2D link length	$[20, 150]$ m
Number of CUEs in cell	30
Number of relays/D2D links	50
Path-Loss model	$128.1 + 37.6 \log_{10} d$
UE/relay Tx power (*fixed*)	20 dBm
Noise power spectral density	-174 dBm/Hz
System bandwidth (BW)	10 MHz

rest of the system parameters are shown in Table 5.1.

We compare the proposed GA techniques (one-point (OP) and two-points (TP) crossover) with the heuristic RA algorithm that was described in Section 4.4 and a random RA method. The random method works as follows: after the allocation of orthogonal RBs to cellular UEs takes place, DUEs are also randomly assigned resources from the available RB pool and satisfy their transmission needs by selecting either relay or direct mode, depending on which of the two modes provides better rate performance.

First, a significant factor that needs to be taken into account is the convergence point of the applied GA methods. This point can be interpreted as the number of generations that results in the optimal achievable aggregate rate, i.e. how fast the algorithm converges to the maximum reached rate with respect to an upper bound of generations and for a defined number of Monte Carlo runs. The box plot in Fig. 4.3 shows that the TP-GA technique converges almost 1.5 times faster compared to the OP-GA (the medians of convergence points in relation to the number of generations are 290 and 412, respectively). Also, the horizontal edges of each box (25 and 75 percentiles) show a bigger gap in the second case where the TP-GA can achieve a really fast convergence on average. This can be

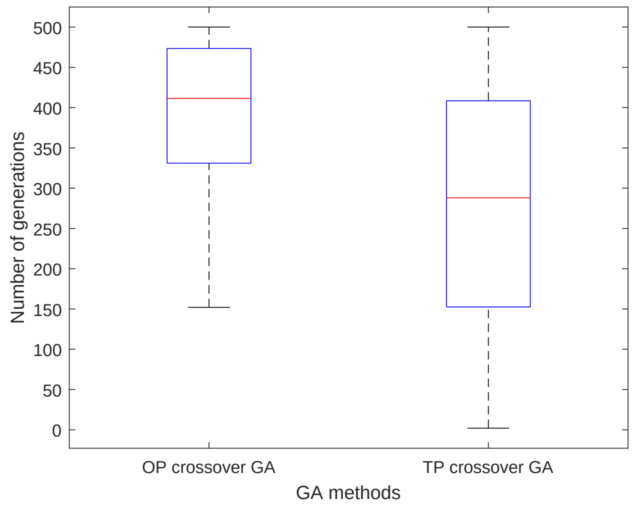


Figure 4.3: Average convergence points for the case of (i) one-point (OP) crossover GA, and (ii) two-points (TP) crossover GA.

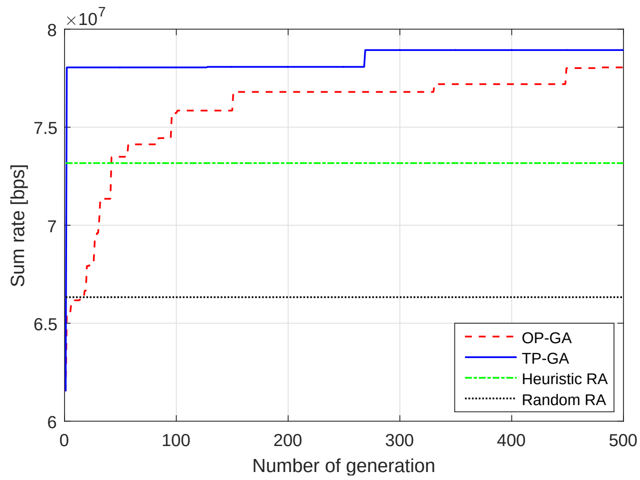


Figure 4.4: An example of the GA's convergence to the maximum rate values.

justified by the TP crossover's ability to ensure a more diverse initial population and encoding that can entail faster convergence to the optimal rate.

Further, Fig. 4.4 shows a sample of the sum-rate performance tendency for a designated number of generations. In this case, the TP-GA not only converges faster to its optimal solution (i.e. 210 generations less needed) but also the achievable rate is notably high compared to the heuristic (almost 10%) and clearly better than the OP-GA method. It has to be noted that, in this case study, the TP-GA method provides a higher capacity performance even from the second generation and beyond, while OP-GA converges in its optimal point in the 468^{th} generation but with rather sub-optimal throughput. Last, for this simulation run, TP-GA outperforms the random method with almost 21% gain in terms of sum rate performance.

Fig. 4.5 illustrates the sum-rate performance of the proposed methods when the D2D transmitter and receiver are separated by fixed distances for each evaluation point. The TP crossover GA method achieves an average sum-rate gain of 4%, 24% and 43% compared to the OP-GA, heuristic and random allocation techniques, respectively. The plot shows that even though the rate drops proportionally with the increase of the D2D link range, the performance gap of the GA proposed algorithms in comparison to the two RA schemes becomes larger. For the case that the D2D link length is fixed at 250 meters for all DUEs, the TP-GA method provides a rate improvement of 37% and 72% compared to the heuristic and the random methods, respectively, signifying a more efficient resource and mode (direct, relayed) selection for D2D communications.

Finally, we investigate the received interference by D2D UEs for all the considered cases. Note that, this interference can result from both a cellular and other D2D/relay transmissions that reuse the same spectrum. As shown in Fig. 4.6, the GA methods achieve a lower interference level where at the 50th percentile, the interference level in GA is 4.7 and 10 dB lower than the heuristic and random

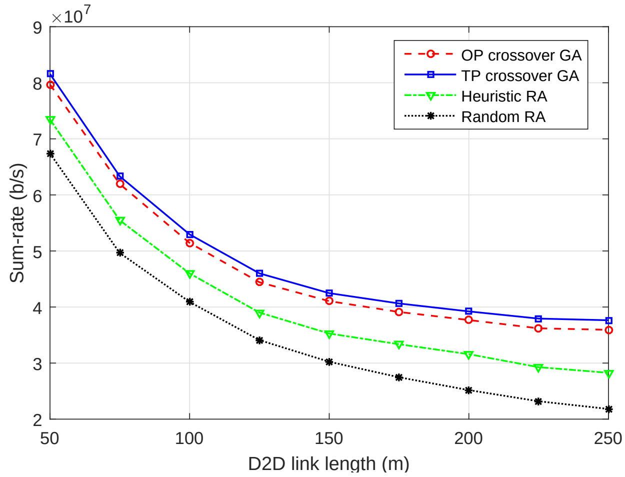


Figure 4.5: Aggregate throughput in relation to varying D2D link lengths.

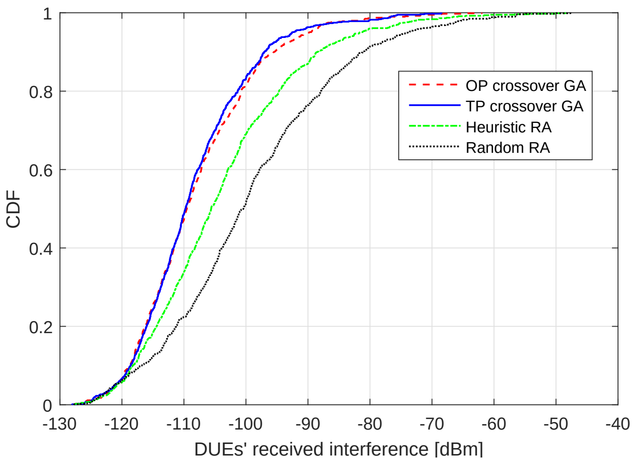


Figure 4.6: CDF of DUEs' received interference.

methods respectively, and at the 90th percentile the GA interference reduction is 9.4 and 15.7 dB compared to the baseline methods.

4.6 Summary

In this chapter, we presented a resource allocation methodology for relay-aided D2D communications that underlay a cellular network. By exploiting the robustness and versatility of bio-inspired meta-heuristic techniques, we proposed a low-complexity genetic algorithmic framework that aimed at maximizing the network throughput performance with respect to interference. Numerical evaluation results highlight the merits of the investigated GA methods. The proposed one-point and two-points crossover GA techniques provide significant rate improvement amounting to more than 20% and 40% compared with heuristic and random RA methods respectively. The proposed GA methods also ensure the least exerted interference towards D2D transmissions with an average gain of more than 4 and 9 dB as compared to the baseline techniques.

Chapter 5

Virtualized Resource Management for D2D Communications in 5G Networks

5.1 Introduction

Network and radio resource management is undergoing a significant change which relates to a number of different underpinning forces. Firstly, to the provision of new and emerging Internet services with increased aggregate volumes of traffic, higher user demands and fast changing requirements. Indicatively, latest Cisco's forecasts envisage an explosive mobile data traffic growth that will reach an almost 10-fold increase by 2020 [1]. Secondly, we are witnessing higher network heterogeneity and the emergence of multiple stakeholders with the overarching need to significantly reduce deployment costs and achieve a sustainable network operation. To this end, network virtualization has recently emerged as a promising technique to overcome the complexity of current network operation as well as facilitate inter-operators sharing [80]. Therefore, efficient approaches to manage radio and network virtualized resources, are expected to be a catalyst element of future mobile network architectures.

Stepping back for a while and taking a more holistic view, it can be stated that, depending on the reused components, network sharing approaches can be classified into active and passive sharing [81]. Active sharing accounts for efficient reuse of key infrastructure components, such as backhaul connections, base stations and, ultimately, the radio access network. On the contrary, passive sharing, which takes place widely today, relates to the cell-site based reusing of its functional components, such as the physical infrastructure, pylons, electrical supply and so forth. The combination of both concepts can not only offer flexibility and potential capital and operational expenses reduction for the network operators, but also a flexible and programmable mobile network.

The above vision is starting to take shape via the move towards network function virtualization (NFV) and software-defined networking (SDN) that provide a formal architectural view on softwarization and cloudification for emerging wireless networks [82][83][84]. On the one hand, NFV enables the implementation of innovative applications without taking into consideration the substrate networks as well as it allows the actual virtualization of some network control functionalities [85] [86]. It further gives the ease to mobile virtual network operators (MVNOs) to turn the attention on resource utilization improvements. On the other hand, SDN is the concept that decouples the control and user planes of mobile network devices and provides the means for simplifying network operability as well as leading to enhanced performance.

5.1.1 Brief Description of Software-Defined Networking and Network Function Virtualization

Capitalizing on emerging technologies such as SDN[1] and NFV[2] provides the network operators with the opportunity to speed up their services' innovation, reduce the overall costs coming from the need to conserve the network equipment and the consumed power for its operation.

Software-Defined Networking (SDN)

SDN is an emerging technology that creates a fertile ground for network operators to easily manage today's rate and bandwidth-demanding applications in a cost-efficient, dynamic and adaptable fashion. The idea is to physically separate (decouple) the two networking planes, that of control and forwarding ones. As a consequence, it enables the network control to become directly programmable and the underlying infrastructure shareable and abstracted for several network services and applications [82]. The control plane functions are running on a centralized network controller that is responsible for routing and forwarding the traffic to network elements which constitute the entities of the separated data plane [5]. This SDN controller has a holistic view of the network nodes and is able to dynamically produce and reconfigure the best flow/routing policies in a heterogeneous network of single or multiple vendors. Furthermore, through the northbound[3] interface of the SDN architecture, application programming interfaces (APIs) enable the communication between network applications and the control layer, leading to an automated management of the network and an easy way for applications to be created, tested and deployed in low time-scales.

[1] www.opennetworking.org

[2] www.etsi.org/technologies-clusters/technologies/nfv

[3] The northbound interface describes the area of protocol-supported communication between the controller and applications or higher layer control programs. Functions of northbound APIs include management solutions for automation and orchestration, and the sharing of actionable data between systems.

The Open Networking Foundation (ONF) [82] has also defined *Open Flow* as the prevailing southbound[4] interface to realize the communication between the centralized network controller and the network devices (both physical and virtual switches and routers) located in the transport network (infrastructure layer). Particularly, it provides the means to access and configure the forwarding plane of a group of physical and/or virtual network nodes, such as switches and routers.

Network Function Virtualization (NFV)

In envisioned future virtualized and programmable 5G wireless network architectures, different tenants (virtual network providers) will be sharing the physical (substrate) network resources using a combination of SDN and NFV architectures. Now, the core idea behind NFV that comes as a complement to SDN technology is to capitalize on virtualization technologies to decouple physical network equipment from the services or functions that run on top of them [85]. Under the NFV framework, a network service can be decomposed into a set of virtual network functions (vNFs) which are implemented in software and are able to run in general purposed hardware where they can be dispatched on demand. An overview of the NFV architecture is shown in figure 5.1 which conforms to the ETSI NFV framework. As depicted in this figure, a service request will be handled by the Orchestrator which will then inform the Virtual Function Manager (VFM) about which vNFs are required to be activated for this specific service, whereas the actual physical resources for the vNFs will be handled by the Virtualization Infrastructure Manager (VIM). The above defined policies for the service creation

[4]The southbound interface is the OpenFlow (or scarcely alternative) protocol specification. Its main function is to enable communication between the SDN controller and the network nodes so that the router can discover network topology, define network flows and implement requests relayed to it via northbound APIs. Functions of southbound APIs include communication with the switch fabric, network virtualization protocols, or the integration of a distributed computing network.

will be distributed using the SDN controller (based for example on OpenFlow[5]).

Since each tenant will be allocated a slice of the available network resources (including also spectrum), mobile users that will require D2D communication from different tenants will be allocated resources (RBs) from the device which is originating the communication. This operation might lead to inefficient usage of the tenant's available resources in the long run. We therefore propose the use of an inter-slice coordinator that will allow for optimal usage of multiple tenants' resources in the case where the communication is taking place between users subscribed to different tenants. An illustrative example of an inter-slice controller is shown in figure 5.8, for the case of two tenants. Such cross-tenant orchestration would allow a more efficient use of the available physical resources per tenant. The concept of inter-slice coordination is being developed within the EU 5G-PPP 5G-NORMA project where the key motivation is to replace single RAN's networked entities by a network slice with a graph of programmable network functions[6]. A cross-tenant controller should be a trusted entity since in order to optimize the overall performance tenants will have to provide intra-slice topological information to the controller which might include, inter alia, the number and location of users in each slice. Depending on the actual implementation, the cross-tenant controller can be considered as a broker that runs by the substrate network provider, which can be deemed as a trusted element. Note, there is an incentive for all tenants to cooperate since overall network performance is increased; however, trusted entities in virtualized architectures is a topic well beyond the scope area of this thesis. Based on that fundamental assumption, the aim of this work is to quantify the potential achievable gains enabled by such a cross-slice controller.

[5]Open Networking Foundation, OpenFlow Switch Specification Version 1.3.2 April 25, 2013

[6]Mark Doll, 5G NORMA "A Novel Radio Multi-service adaptive network Architecture for the 5G era", 1^{st} *Sino-Europe 5G Workshop*, November 2015, Beijing, China

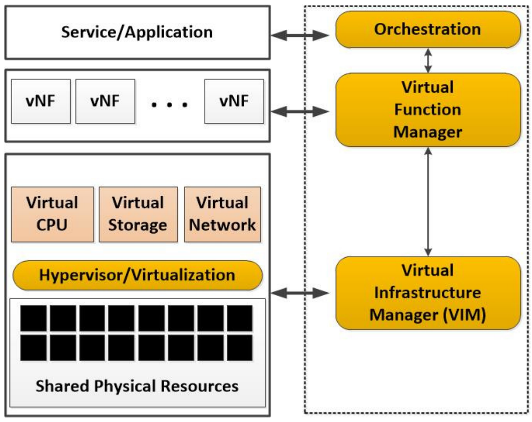


Figure 5.1: Network Function Virtualization architecture following the ETSI framework.

5.2 Optimal Virtualized Resource Sharing for D2D Communications

5.2.1 Contribution

Despite the fact that a number of solutions for RAN virtualization emerged over the last few years, it is worth pointing out that little attention has been placed on issues related to D2D virtualization. In essence, this section shortens the gap between these two important areas by proposing a set of optimization problem formulations to extend previous works on RAN virtualization and explicitly provide resource slicing for D2D communications. To this end, we aim to yield upper bounds on network performance by devising optimal D2D resource slicing via mathematical programming formulations. In addition, sub-optimal low complexity algorithms, amenable to practical (real-time) implementation, are detailed. Via a wide set of numerical investigations, we show that the proposed solution achieves significant gains in terms of system throughput compared to previous related resource slicing techniques which are D2D oblivious.

5.2.2 Closely Related Work

As already mentioned, resource allocation in the context of D2D communications for emerging and future wireless networks is one of the active areas of research [87]. The literature, regarding the integration of D2D communications in cellular networks as well as the orchestration of the scarce network's radio resources, is rather rich and includes a number of different approaches and novel techniques. Optimal utilization of RBs in LTE-A networks is well known to be an NP-hard optimization problem, with or without the presence of D2D links. To this end, a common route has been to propose sub-optimal algorithms for resource utilization using graph-theoretic approaches [88] or by relaxing some of the constraints (e.g. power constraints or the integrality of the resource block allocation) and propose near-optimal heuristic methods for D2D resource allocation [89].

Considering the virtualization of RAN resources, one of the most important functional entities is that of the hypervisor, which is essentially a virtual resource controller that slices the entities of a physical network into different virtual networks. The communication modes are then defined by a central software controller that runs on top of the virtualized infrastructure providers (InPs) and, consequently, the integration of different communication types can be eased. Further detailed information on the adaptation of D2D communications' peer discovery and resource management can be found in [90].

A number of previous closely related works considered the exploitation of NFV and SDN concepts on radio access networks. Notably, Soft-Cell [91] focuses on redesigning the core mobile network under an SDN framework, where the emphasis is placed on the realization of adaptive traffic policies for user data traffic across the core and wireless access network. On a parallel effort, SoftRAN [92] focuses on the radio access network by considering a logically centralized control plane for allocating radio resources. The idea is that by providing a multi-cell view on the network, radio resources can be managed in a more coordinated manner;

hence, SoftRAN, focuses on multi-cell resource allocation whereas in this section we are focusing on D2D resource allocation within a single cell that gives an intuition of the overall network point of view. For a more meticulous overview and of tutorial-style analysis of current efforts on wireless network programmability via NFV/SDN approaches, we refer the interested reader in [93].

As already alluded above, the integration and proper orchestration of D2D in virtualized SDN-based cellular networks is expected to become an important topic for providing low-cost network operation via the virtualization of the underlying functional blocks as pertain to the issue of D2D resource allocation. One of the first efforts towards this direction is the work in [90], where the authors address the problem of network state information (NSI) imperfectness in virtual wireless networks and resource allocation for the software defined D2Ds. They devise a discrete stochastic optimization formulation to the problem of resource sharing given imperfect NSI and, then, proceed with the introduction of stochastic approximation algorithms for both static and varying channels resource manageability. In contrast, in this section we consider the exploitation of an aggregated radio resource pool among multiple MVNOs and D2D communications in a virtual wireless network to maximize the network-wide welfare. The key contribution is the proposal of a D2D-explicit virtualized resource sharing methodology that, compared to other works, makes full use of the available resources from different slices to optimize the performance of D2D users in terms of sum-rate and at the same time retain the interference in acceptable levels. We provide low complexity heuristic algorithms as well as upper bounds on the performance by formulating integer mathematical programming models that also allow the allocation of contiguous RBs to D2D users in order to enhance their aggregate performance.

5.2.3 System Model

We consider a D2D underlay-based cellular network where D2D links and cellular users are randomly distributed in a hexagonal cell layout [12]. The cell's centre-located base stations are considered to be equipped with omni-directional antennas and are being shared by $\mathcal{N}$ InPs. We focus on the most popular UL based resource sharing case, where D2D communications reuse the cellular spectrum of an LTE-A system. The uplink resource use, which implies a set of available RBs ($\mathcal{K}$), is studied due to the widely accepted assumption that uplink will be less congested than downlink. During the UL phase, interference from the D2D transmitter to the BS is taking place, hence monitoring of the interference at the base stations should be considered. On the other hand, D2D receivers are also exposed to interference from the cellular transmissions whose resource blocks are being shared. The difference of the proposed technique compared with the network virtualization substrate (NVS) technique, as presented in [80], is explained in Figure 5.2. This figure shows an NVS-based allocation, where the available resources are sliced for different InPs (slices) to serve their corresponding users' transmissions. However, it does not consider explicitly the D2D communication links. In that case, a RB can be re-used between a CU and a D2D link. As shown in this figure in slice 2, the $D2D_2$ pair reuses the RB allocated to CU_2, resulting in high interference that leads to sub-optimal performance. On the contrary, while the proposed technique implements resource allocation that regards the D2D links of multiple slices, the $D2D_2$ pair would be preferably assigned with a RB from slice 1.

Furthermore, we proceed with the system definition in order to formally discuss the proposed scheme. Without loss of generality, we consider two MVNOs or service providers that acquire and utilize radio resources from different InPs ($\mathcal{N}$). This consideration deviates from the conventional resource allocation case where the cellular spectrum is available for all users (Figure 5.3(a)). As further

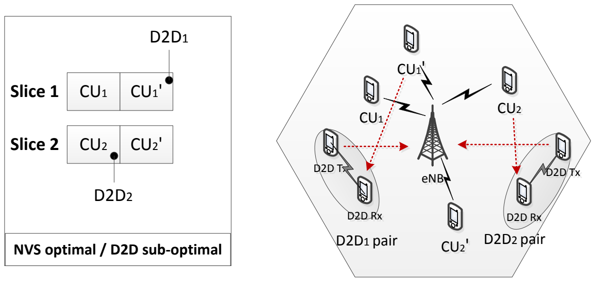


Figure 5.2: An example that shows an NVS-compatible sub-optimal resource assignment case. Dashed lines represent the existed interference.

illustrated in Figure 5.3, the proposed technique can be deemed as an extension of the NVS (case (b)) where a layer of D2D-specific resource assignment procedure is embedded. Our contribution can be represented by the sub-figure 5.3(c) where a functional D2D virtualization block is integrated to fuse the D2D available resource pools of the two slices. In essence, D2D links are enabled to be allocated radio resources from the slice that they do not belong to in contrast to NVS rationale that imposes the allocation of resources from the slice manager that the users are subscribed to. Thus, efficient sharing of multiple slices' resources can be applied in order to facilitate the transmission of D2D UEs and effectively avoid any potential performance degradation due to the spectrum reuse.

In our system model, the path-loss is calculated as follows,

$$PL_{D2D} = 148 + 40 \log_{10} d \tag{5.1}$$

$$PL_{CU} = 128.1 + 37.6 \log_{10} d \tag{5.2}$$

for D2D pairs and cellular users, respectively [94][95]. Parameter d stands for the distance and is expressed in km. In order to present a mathematical programming

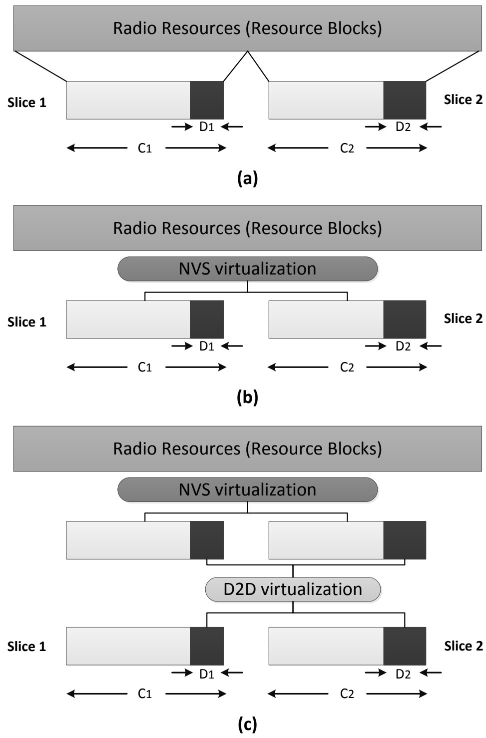


Figure 5.3: Proposed RB availability pool for D2D links for a number of InPs and slices (in this case $|\mathcal{N}|= 1$ and 2 slices are implied). $\mathcal{C}_i$ and $\mathcal{L}_i$ stand for the corresponding resource pools for cellular and D2D links served by slice i.

framework, we define the following binary decision variable:

$$y_{lnk} = \begin{cases} 1, & \text{if D2D link } l \text{ of InP } n \text{ utilizes RB } k \\ \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

Also, we denote with $\mathcal{L}$ the set of D2D links, $\mathcal{N}$ is the set of InPs and $\mathcal{K}$ is the aggregate set of the available resources.

In order for a RB $k = k_l$ to be allocated to a D2D link l of InP n, the required SINR threshold (γ_t) needs to be satisfied to the receiver (for an LTE based network, we consider that this value should be above 7 dB when translated from corresponding SINR ratio values in order to ensure a fair signal and QoS reception). This practically can be expressed as follows,

$$\gamma_{lnk_l} = \frac{g_{lnk_l} P_l}{\sum_{k \in \mathcal{K}} y_{lnk} g_{nk}^{cl} P_c + \sigma^2 + I} \geq \gamma_t \tag{5.4}$$

where P_l is the transmission power of the D2D transmitter of link l, g_{lnk} is the link gain of the l^{th} D2D link from InP n when using the RB k. g_{nk}^{cl} expresses the link gain between CU c and the receiver of D2D link l of InP n when using RB k (CU-D2D interference will be developed when $k = k_l$ and $y_{lnk} = 1$ in the denominator). Lastly, σ^2 denotes the lump sum power of background/thermal noise and Q the co-channel interference from other cells (if existent).

Also, the SINR threshold ($\tilde{\gamma_t}$) needs to be satisfied also for the cellular transmissions that utilize the same RB (e.g. $k = k_c$) with a D2D link l when transmitting in the uplink period (as before, acceptable value is 7 dB and above). This SINR constraint can be written as follows,

$$\frac{g_{nk_c}^{cb} P_c}{\sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} y_{lnk} g_{nk}^{lb} P_l + \sigma^2 + Q} \geq \tilde{\gamma_t} \tag{5.5}$$

where P_c is the transmission power of a cellular user, g_{nk}^{cb} is the link gain between the CU c and its associated BS b when using RB k, whereas g_{nk}^{lb} accounts for the link gain between the D2D transmitter of link l and the BS b that transmit/receive in the same k^{th} channel.

Considering the above, the achieved rate for D2D link l's receiver of InP n that utilizes the resource block k can be estimated through the Shannon capacity formula accordingly,

$$R_{ink} = B_{RB} \log_2 (1 + \gamma_{lnk}) \tag{5.6}$$

where B_{RB} is the LTE-based resource block bandwidth (180 kHz) and γ_{lnk} is expressed in power ratio. Note also that, in the aforementioned formulas, link gains incorporate also channel fading and shadowing impairments. A shadowing standard deviation of 8 dB for both CU and D2D users is taken into consideration.

In the following two subsections, a set of optimization problems is proposed for maximizing sum-rate performance of D2D pairs under a given number of available resources. Then, a heuristic resource slicing algorithm is devised to provide a low complexity albeit sub-optimal performance.

5.2.4 Single Resource Block Sharing in Virtualized Environments

In this subsection, we turn the focus on the allocation of a single LTE RB as the minimum assignable resource unit for each deployed D2D link. We assume that all cellular users (we denote this set with $\mathcal{C}$) are being allocated with orthogonal resources and reserve a single RB to satisfy their transmission needs (each c is assigned different RB k_c). We subsequently devise an optimization problem to maximize the sum-rate of the involved D2D users in a manner that the overall CUs' performance is not critically affected by the potential reuse of their resources. The mathematical setting can be developed as follows,

$$\max \sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} R_{lnk}y_{lnk} \tag{5.7}$$

subject to:

$$\sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} y_{lnk}g^{lb}_{nk}P_l\tilde{\gamma}_t \leq -\Big(\tilde{\gamma}_t(\sigma^2+Q) - g^{cb}_{nk_c}P_c\Big),$$

$$\forall c \in \mathcal{C},\ k_c \in \mathcal{K} \tag{5.7a}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} y_{lnk} = 1,\ \forall l \in \mathcal{L} \tag{5.7b}$$

$$\sum_{l\in\mathcal{L}}\sum_{k\in\mathcal{L}} y_{lnk} \leq U_n,\ \forall n \in \mathcal{N} \tag{5.7c}$$

$$y_{lnk} \in \{0,1\},\ \forall l \in \mathcal{L},\ \forall n \in \mathcal{N},\ \forall k \in \mathcal{K} \tag{5.7d}$$

where U_n is the maximum number of users that can be served by InP n and $\tilde{\gamma}_t$ stands for the SINR threshold that needs to be satisfied for a successful UL cellular transmission. Constraint (5.7a) stands for the satisfaction of the SINR requirement for the cellular transmissions and is derived from equation (5.5), whereas (5.7b) represents the assignment of only one RB per D2D link l. Finally, constraint (5.7c) represents the limitation in terms of usable resources per InP.

5.2.5 Multiple Resource Block Sharing in Virtualized Environments

We augment the previously defined formulation to allow for multiple resource block allocation per D2D link. In the uplink RB allocation, a constraint that needs to be satisfied is that, if more than one RBs are allocated to a user, these should be contiguous according to the SC-FDMA requirements [96]. In order to apply a multiple-RB allocation optimization framework for D2D users, an additive decision variable needs to be defined:

$$s_{lnk}^{w} = \begin{cases} 1, & \text{if } \{k, \cdots, k+w-1\} \text{ RBs allocated to } l \\ \\ 0, & \text{otherwise.} \end{cases} \tag{5.8}$$

This variable indicates the link l's allocation with contiguous RBs, where the allocation's starting point within a block of resources is k and expands to w in total consecutive positions. For example, if link l of n utilizes 2 RBs starting from resource block identifier with $k = 3$, this can be represented as $s_{ln3}^{2} = 1$. Assuming an upper limit of Γ consecutive RBs allocable per D2D pair, we empower the optimization program mentioned before as follows:

$$\max \sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} R_{lnk}y_{lnk} \tag{5.9}$$

subject to:

$$\sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} y_{lnk}g^{lb}_{nk}P_l\tilde{\gamma}_t \leq -\Big(\tilde{\gamma}_t(\sigma^2+Q) - g^{cb}_{nk_c}P_c\Big),$$

$$\forall c \in \mathcal{C},\ k_c \in \mathcal{K} \tag{5.9a}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} y_{lnk} \geq 1,\ \forall l \in \mathcal{L} \tag{5.9b}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} y_{lnk} \leq \Gamma,\ \forall l \in \mathcal{L} \tag{5.9c}$$

$$\sum_{l\in\mathcal{L}}\sum_{k\in\mathcal{K}} y_{lnk} \leq U_n,\ \forall n \in \mathcal{N} \tag{5.9d}$$

$$y_{ln(k-1)} - y_{lnk} + y_{lnm} \leq 1,\ \forall m \in \{k+1,\ldots,|\mathcal{K}|\} \tag{5.9e}$$

$$y_{lnk} + y_{ln(k+m)} \leq 1,\ \forall m \in \{\Gamma,\ldots,|\mathcal{K}|\} \tag{5.9f}$$

$$\sum_{m\in M} y_{lnm} \geq l \cdot s^w_{lnk},\ M = \{k,\cdots,k+w-1\} \tag{5.9g}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{w\in\mathcal{W}} s^w_{lnk} = 1,\ \forall l \in \mathcal{L} \tag{5.9h}$$

$$y_{lnk} \in \{0,1\},\ \forall l \in \mathcal{L},\ \forall n \in \mathcal{N},\ \forall k \in \mathcal{K} \tag{5.9i}$$

$$s^w_{lnk} \in \{0,1\},\ \forall l \in \mathcal{L},\ \forall n \in \mathcal{N},\ \forall k \in \mathcal{K},\ \forall w \in \mathcal{W} \tag{5.9j}$$

where $\mathcal{W}$ is the set $\{1,\ldots,\Gamma\}$ and depends on the initialization of Γ when the aggregate resource pool can be further utilized. Considering the added constraints, (5.9c) presents the maximum assignable number of consecutive RBs for a D2D link l. Constraint (5.9e) accounts for excluding the case of an unallocated RB ($y_{lnk} = 0$) between two (or more) assigned (e.g. $y_{ln(k-1)} = 1$ and $y_{ln(k+1)} = 1$) to the same user, whereas (5.9f) can be interpreted as the restriction of not having assigned resource blocks after $k+\Gamma$ positions when a link is assigned with k^{th} RB. While with the two latter constraints we ensure the compliance of not having a zero between two ones, on the other hand, constraint (5.9g) is the one to decide for the consecutiveness of the RB assignment to a D2D link l; w signifies the number of successive resources to be assigned to the link if the assignment starts from RB k. Furthermore, sub-equation (5.9h) declares that only one of the s_{lnk}^{w} components for each user should be satisfied (i.e. valued with 1).

The $\mathbf{s}$ decision variable can be then visualized in a vector form,

$$\mathbf{s} = \begin{bmatrix} \mathbf{s_1}, & \mathbf{s_2}, & \cdots & \mathbf{s_{|\mathcal{L}|-1}}, & \mathbf{s_{|\mathcal{L}|}} \end{bmatrix}^{\intercal}, \tag{5.10}$$

where $\mathbf{s_l}$ corresponds to each D2D link $l \in \mathcal{L}$. If we denote with R_{tot} the total number of available resources, the $\mathbf{s}$ variable vector's length is $|\mathcal{L}| \cdot \sum_{m=1}^{\Gamma} (R_{tot} - m + 1)$, the Euclidean norm for each link l is $\|\mathbf{s_l}\| = 1$ (according to (5.9h)) and fluctuates as it is Γ-dependant. Finally, (5.9i) and (5.9j) are the integer boundaries for both decision variables of the problem. The final form of the decision variable for the problem of multiple-RB sum-rate maximization problem is $\mathbf{x} = [\mathbf{y}; \mathbf{s}]$. However, only the integer components of $\mathbf{y}$ participate in the objective function maximization.

5.2.6 Heuristic Algorithm

Algorithm 4 provides in detail an alternative, less complex solution to allocate up to Γ resources per D2D link $l \in \mathcal{L}$. Its allocation rationale is based on initially assigning RBs that provide the best channel conditions to each respective user. The algorithm runs sequentially for the number of involved slices in order to retain a fair assignment of the maximum SINR-based sorted resources for all D2D users. Then, and due to the investigation of UL transmission instance that is characterized by the consecutive resource allocation for all users, the algorithm searches for an adjacent RB that provides further performance improvement (i.e. better SINR) for each involved link. It is highly important to notice that in

Algorithm 4: *HRS*

Input : CU-D2D UEs' location coordinates, Γ.
$*$ **Initial state** :
- Each $c \in \mathcal{C}$ has already been allocated with a RB.
- Each $c \in \mathcal{C}$ is served by a specific InP $n \in \mathcal{N}$.

$*$ **D2D resource allocation steps** :
{Step 1 : *D2D resource assignment initialization*}
$l = 1$; ▷ D2D link id.
while $l \leq |\mathcal{L}|$ **do**
  if *($l \mod 2 \equiv 1$)* **then**
    • $slice = 1$; ▷ slice id.
    • $l = l_1$; ▷ D2D pair of slice 1.
    • $\forall$ available RBs, find the D2D-RB combination (l_1 user and k_l RB) that results in the maximum estimated SINR among all;
    • Allocate k_l RB to user $l_1 \in \mathcal{L}$;
    • $p(l_1) = \{k_l\}$; ▷ set that keeps l's used RBs.
    • $l = l + 1$;
  else
    • $slice = 2$;
    • $l = l_2$;
    • Repeat the same procedure for $l = l_2$;
    • $p(l_2) = \{k_l\}$;
    • $l = l + 1$;
  end
end
* ($l \mod 2$) ensures the sequential RB allocation of D2D users for the example of two MVNOs/slices.

{Step 2 : *Max-SINR based contiguous RB allocation*}
for $l := 1$ *to* $|\mathcal{L}|$ **do**
  • $p(l) = \{k_l\}$; ▷ recall associated RB for user l.
  • $\rho_l^- = k_l - 1$, $\rho_l^+ = k_l + 1$;
  while $n(p(l)) \leq \Gamma$ **do**
    if *(ρ_l^- gives better channel conditions to link l & is not occupied & $SINR_{c,\rho_l^-} \geq \tilde{\gamma}_t$)* **then**
      • Allocate RB ρ_l^- to user $l \in \mathcal{L}$;
      • $p(l) = \{p(l), \rho_l^-\}$;
      • $\rho_l^- = \rho_l^- - 1$; ▷ continue left-wise.
    else if *(ρ_l^+ satisfies all constraints)* **then**
      • Allocate RB ρ_l^+ to user $l \in \mathcal{L}$;
      • $p(l) = \{p(l), \rho_l^+\}$;
      • $\rho_l^+ = \rho_l^+ + 1$; ▷ continue right-wise.
    else if *(conditions not satisfied)* **then**
      • Break and check next user l;
    end
  end
  • Calculate R_l; ▷ rate of link l (Shannon-based, eq. (5.16)).
end
** $n(p(l))$ stands for the cardinality of set $p(l)$ for D2D link l.

{Step 3 : *HRS aggregate throughput estimation*}
• $SR = \sum_{l \in \mathcal{L}} R_l$; ▷ sum-rate for all $l \in \mathcal{L}$.

order for this RB to be allocated to a specific D2D link, it must be firstly an unallocated one and, secondly, do not degrade the performance of the cellular user that utilizes this resource. The RB assignment continues until one of the following conditions is met: (i) whether the user reaches its RB use upper limit (i.e. maximum Γ RBs can be allocated to it) or, (ii) all resources are utilized and hence the pool of resources is depleted. The algorithm executes this iterative search bi-directionally, thus the finally assigned RBs can be either on the left or the right or in both sides of the initially allocated to D2D pair RB.

5.2.7 Performance Evaluation

Herein, a set of numerical results is being detailed to highlight the expected efficiency of the proposed optimization framework that considers virtualized D2D resource sharing. We consider a single hexagonal cell where D2Ds and CUs are randomly distributed in space. Also, for ease of understanding, a single InP is taken into account but the proposed technique can be readily applied for multiple InPs. A summary of the related system parameters is shown in Table 5.1. Evaluation results derive from Matlab-based Monte Carlo simulations.

We compare our optimal resource slicing (ORS) method with a number of sub-optimal schemes. First, as detailed before, with the proposed heuristic resource slicing (HRS) scheme that provides a sub-optimal performance for the problem of sum-rate maximization in D2D-based cellular network, under the scope of resource virtualization. Second, we resemble the work in [80] and implement it with respect to the D2D communication concept by slicing radio resources according to the NVS resource-based provisioning scheme (for the rest of the text we call this *NVS-A* scheme); specifically, it dedicates a fraction of the total BS resources for each slice to serve its corresponding D2D users. Compared to our scheme, we take into consideration a full-sharing RB pool by fusing the D2D-available resources for all involved D2D links. In order to apply fair comparison

Table 5.1: Simulation Parameters

Parameter	**Value**
Cell layout	Hexagonal grid
CU-D2Ds distribution	Uniform
Macro cell radius	400 m
Maximum D2D link range	50 m
Number of CUs in cell ($\lvert\mathcal{C}\rvert$)	30
D2D Path-Loss model	$148 + 40 \log_{10} d$
CU-BS Path-Loss model	$128.1 + 37.6 \log_{10} d$
Maximum D2D Tx power (P_l)	15 dBm
Maximum CU Tx power (P_c)	20 dBm
Shadowing standard deviation	8 dB
Noise power spectral density	-174 dBm/Hz
System bandwidth (BW)	10 MHz

between this technique and the *HRS* proposal, *NVS-A* bases its rationale on the allocation of the best RB to the best D2D link per slice in a sequential manner, until all users are served. Third, noted as *NVS-B*, is a pre-emptive method that runs sequentially for each slice and randomly for an associated D2D link. It allocates the best SINR-providing RB and extracts this resource from the RB available pool. Last, and as a baseline, we have implemented the *NVS-C* scheme which randomly allocates a RB from the corresponding resource pool to a randomly picked D2D user to serve its transmission needs. It is important to note that all aforementioned methods take into account CU performance, i.e., not to violate the corresponding SINR threshold. In addition, for the multiple-RB allocation problem and after the initial assignment of one RB, the bi-directional consecutive RB assignment is triggered to make efficient use of the resources.

Regarding the problem of single-RB allocation formulated in (5.7), we compare the performance of the schemes mentioned before in terms of D2D mean sum-rate (Figure 5.4). As expected, the *HRS* algorithm gives a near-optimal solution as it is on average only 0.35% less efficient compared to the ORS formula (5.7) and their graphics are nearly tangent. The *ORS* gain becomes more explicit in higher D2D traffic scenarios (20 D2D links in total). Additionally, the optimal solution outperforms the *NVS-A* technique in a mean percentage of almost 6.3%, whereas the gap increases for the case of *NVS-B*, where a D2D sum-rate differentiation of over 8.5% is observed. The increasing tendency of the performance of the ORS and HRS solutions becomes more clear in the case of more congested from D2D users scenarios. Finally, the random allocation NVS-based scheme (i.e. *NVS-C*) results in the worst among all performance with an average sum-rate deterioration of almost 116% compared to the optimization solver.

The performance improvement becomes more observable when the ORS is applied for multiple-RB assignment scenarios, as presented in Figure 5.5. Based on the same varying cases of D2D traffic and for a maximum allocable number of four RBs (i.e. $\Gamma = 4$), the heuristic algorithm behaves better compared to the NVS-based methods. According to the depicted results, the optimal *ORS* solution overrides the *HRS* algorithm in a mean percentage of almost 15%, while the latter seems to converge to the optimal solution in more bottlenecked scenarios; for a number of 20 D2D links, the *HRS*-based sum-rate is approximately 4% worse than the optimal proposal. Considering the rest of the techniques, the mean sum-rate performance improvement of the optimization proposal is nearly 27%, 33% and 89% compared to *NVS-A*, *NVS-B* and *NVS-C*, respectively.

Figure 5.6 indicates the normalized RB utilization percentage of the compared schemes. It can be easily understandable that the optimization problem will utilize the maximum number of available resources to maximize its objective throughput performance. However, there might be the case where the limited resources need to not be overly utilized to avoid any resource deficiency and inability to serve future transmissions. The main deduction point of this figure is that *HRS* can achieve competitive sum-rate performance by making at the same

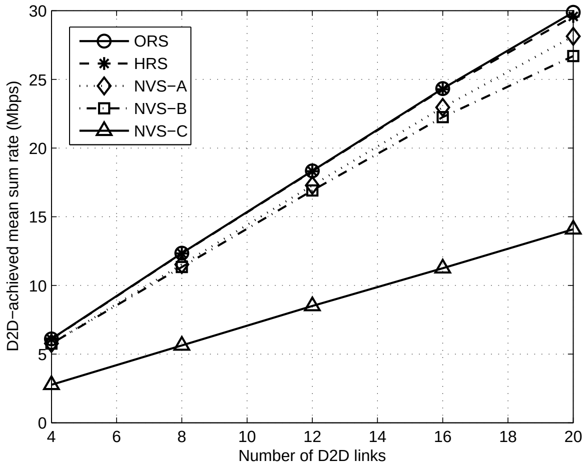


Figure 5.4: Mean sum-rate estimation for varying number of D2D links. A single RB is allocated for each D2D link for this realization. Maximum link range of 50 meters is assumed in this study.

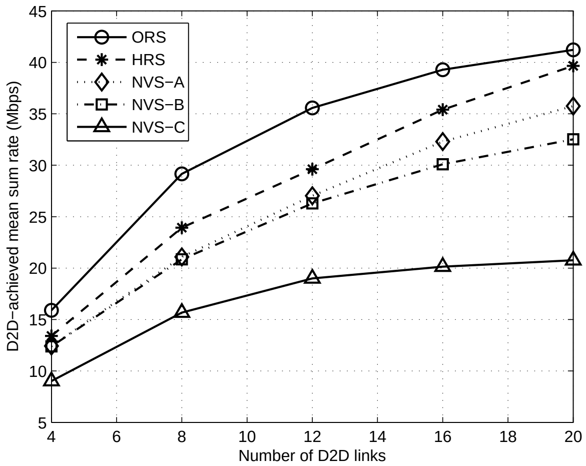


Figure 5.5: Sum-rate estimation for varying number of D2D links. Maximum bound of $\Gamma = 4$ is set for this realization.

time use of the least resources compared to the rest of the approaches. More specific, in light D2D traffic scenarios, the *HRS* exploits the resource allocation rationale of assigning first the best available resources of the fused D2D RB pool and achieves the least RB utilization. As the traffic increases, it is obvious that more RBs are being used and the utilization levels' gap becomes shortened for all of the investigated techniques.

Lastly, we performed an analysis on the effect of the resource allocation schemes to the transmission of cellular users in the UL. The results derive from a number of simulations where the ratio of D2D links with the cellular users is over 50% to indicate a high traffic scenario. The metric used for the realization of this comparison is the received signal to the BS, which, due to the interference from D2D transmitters to the cellular transmissions, degrades accordingly. Figure 5.7 depicts the average BS received signal that for both traffic scenarios highlights the supremacy of the *ORS* solution in terms of least interference. *ORS* and *HRS* methods outperform the NVS techniques by achieving minimum interference to the BS when the number of D2D is increased and as shown in the previous paragraph when the resources used are arithmetically similar. The differences in terms of received signal are notable but also slight due to the low transmission power of the interfering D2Ds. Compared to the baseline (i.e. *NVS-C*), *ORS* prevails in a mean percentage of almost 23% for both studied instances.

5.3 Optimal Inter-Tenant Resource Sharing for D2D Communications

5.3.1 Contribution and Structure

As discussed, the virtualization of wireless radio resources has arisen as a promising solution to encounter the ongoing increasing data demand in today's and

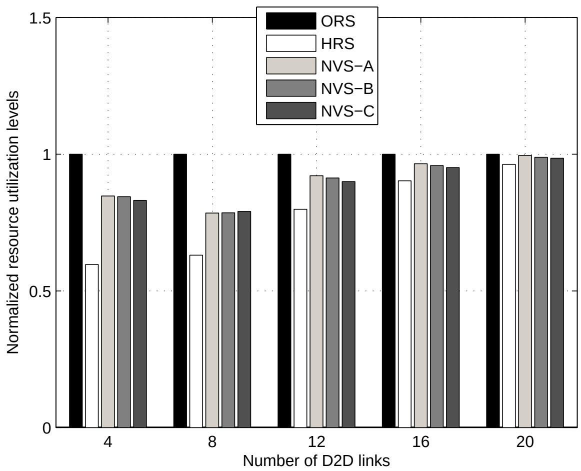


Figure 5.6: Normalized resource block utilization levels for multiple-RB sharing.

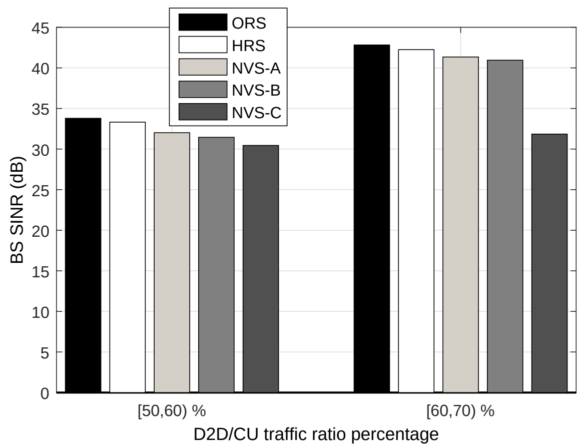


Figure 5.7: BS SINR in UL in relation to high traffic congestion levels and multiple D2D pairs.

emerging future networks. Wireless resource virtualization (WRV) is currently emerging as a disruptive technology that offers significant benefits to different networks and service providers (N&SPs) [97] as well as enabling vertical industries to create their own wireless network. Briefly, other than the fact that co-existing networks, which will be called tenants in the sequel, are able to share the substrate physical infrastructure that entails reduced capital and operational expenditures, WRV ameliorates the utilization of radio resources via sharing them among the different N&SPs [98]. Hence, the exploitation of the wireless virtualization and network programmability merits on top of the integration of D2D paradigm can lead to improved network performance in terms of spectrum efficiency as well as overall network performance in future wireless ecosystems.

However, the advent of the data-driven era brings in a number of challenges and complexity mainly due to the resulting cell and user densification. Among all, a prevalent problem that is expected to attract not only academic but also industrial interest is that of direct communication between users that are subscribed to different mobile network operators (MNOs). A solution to this problem could create a fertile ground for introducing new business models that will fully leverage the D2D potentials. Technically, the weight should be primarily put on defining how the involved MNOs will coordinate their spectrum to satisfy their subscribers' QoS requirements. In the case of single MNO, underlaying D2D links are allowed to utilize the licensed cellular spectrum that is provided by the operator. On the other hand, in the case where two devices belong to two different operators, it needs to be decided which resources from which MNO will be utilized to realize the D2D connection. Therefore, the principal aim is to support a significant number of direct connections along different network operators while at the same time respect the performance of cellular users as well as the overall welfare of the system.

In this section, we assume that separate network slices allow for multi-tenant

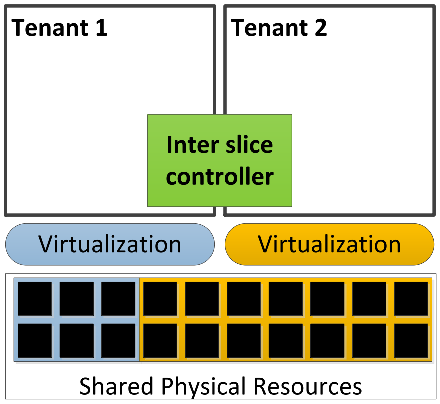


Figure 5.8: A cross tenant communication entity that would allow efficient use of resources for users belonging to different tenants.

D2D discovery and session initiation procedures with similar techniques as the ones defined in [99], where a protocol is designed to permit the inter-operator D2D communication. Without loss of generality, we consider two tenants and for each tenant we assume a number of subscribed CUEs and DUEs, randomly distributed in a typical hexagonal cell layout. Each DUE is considered to be communicating in a maximum allowed distance with a peer that belongs to different tenant. Based on this topological modeling, we propose an ILP optimization framework that aims at maximizing the sum rate performance of the involved inter-tenant D2D links while retaining the cellular UEs' QoS requirements of the involved tenants above a predefined performance threshold. To the best of our knowledge, this is the first work that deals with the inter-tenant D2D communication optimization in virtualization-enabled networks.

5.3.2 Closely Related Work

So far, the problem of inter-operator D2D communication is inadequately explored, thus it needs to be carefully encountered in order to harvest the business dynamics of this specific communication type. The only existing work on this particular topic is [100]. Therein, the authors propose the allocation of inter-operator D2D communications over dedicated licensed radio resources (overlay D2D) which the different operators have to negotiate between each other about the amount of spectrum that they will finally dedicate. They formulate this problem as a game between two distinct mobile network operators and decide about the offered dedicate spectrum with a best response method that runs in a sequential manner. Compared to it, our approach differs in that we consider a virtualized RAN infrastructure where inter-operator D2D links can utilize the whole available spectrum (underlay as opposed to overlay) to achieve efficient resource sharing other than sum-rate maximization.

However, even the dynamics of intra-operator D2D communications in RAN virtualized ecosystems are barely explored in up-to-date literature. Recalling form the latest section, one of the first efforts towards this direction is the work in [90], where the authors address the problem of NSI imperfectness in virtual wireless networks and resource allocation for the software-defined D2D connections. They devise a discrete stochastic optimization formulation to the problem of resource sharing given imperfect NSI and, then, proceed with the introduction of stochastic approximation algorithms for both static and varying channels resource manageability. Further, our work in [101] considered the virtualization of the resources offered by different MVNOs in order to support and improve the performance of intra-MVNO D2D connections in the uplink scenario. The problem was formulated as an ILP sum-rate maximization problem, based on the constraint that the allocated resources per D2D must be contiguous. Heuristic proposed methods were also included as low-complexity solutions.

Virtualization of the core as well as the radio access network is envisioned as the de-facto way forward for 5G networks since it can provide higher degree of flexibility to the mobile network operator, whilst with a careful design it can reduce overall network cost [80],[102]. A preliminary study as with respect to use cases and requirements has also been defined within the 3GPP [103]. Also, important architectural aspects have also been discussed in order to support such advanced mechanisms [104]. Under the assumption of a virtualized mobile network, the work in [105] considers the issue of resource allocation for D2D nodes via a non-linear optimization framework but does not consider the issue of inter-tenant D2D resource optimization.

5.3.3 Problem Definition

Preliminaries

The studied hexagonal cell area consists of a center-located RAN virtualized BS equipped with omni-directional antennas and a number of cellular users and D2D links uniformly distributed. Part of the distributed cellular and D2D users are assumed to belong to a specific tenant and are served by its designated slice, whereas the rest of them are subscribed to a second tenant, hence, a separate slice is dedicated to serve them. Note that, without loss of generality, we hereafter assume the existence of two tenants. Figure 5.9 depicts the described scenario where intra and inter-slice/tenant D2D communications can take place. The setup of D2D communication is out of the scope of this thesis. Briefly, intra-operator D2D session setup is carried out by the session initiation protocol (SIP) discussed in [12], whereas the establishment and realization of the inter-operator D2D connection is detailed in [99].

As already mentioned in the Introduction, each tenant is assigned with a slice that will provide, inter alia, spectrum allocation in order to fulfill the expected demand from the serving users. Quantitative, this translates to a number of RBs which constitute the available resource pool of the users that are subscribed to specific tenant. Considering the legacy procedure, in order to support a D2D link between two users (intra or inter-operator), the resources used for it are allocated only from the RB pool that corresponds to the user that inaugurates the direct communication. However, in this work, we leverage the ability of an inter-slice

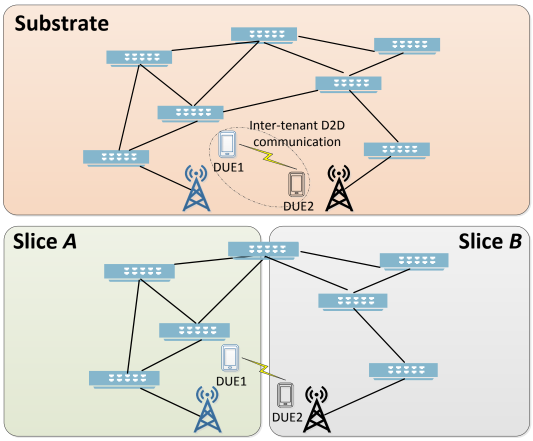


Figure 5.9: Illustration of network slicing and inter-tenant D2D communication.

manager to fuse the available RB pools of multiple tenants in order to enlarge the spectrum availability for inter-operator D2D links (figure 5.8).

Following the principles of D2D cellular spectrum reusability in the underlay notion, D2D users are able to utilize the resources of multiple cellular users simultaneously [106]. However, it is important to ensure that the cellular transmissions whose resources are being reused by a D2D pair satisfy their QoS minimum requirements. For ease of comprehension, we presume that cellular users primarily occupy one but orthogonal radio resource based on the LTE specifications.

Problem Formulation

The reason why we consider this multi-tenant unified RB pool is to increase the resource efficiency for cross-tenant D2D links which are expected to be a significant part of future network connections. This will not only lead to effective usage of the available spectrum, but also improve the overall network performance by potentially increasing throughput and reducing interference. To this direction, an ILP optimization solution is proposed to maximize the sum-rate for inter-slice D2D links by respecting at the same time the cellular transmissions' performance not to degrade below a predefined threshold.

Before we formulate the D2D sum-rate optimization problem, the following sets need to be defined:

- $\mathcal{L}$ is the set of distributed cross-tenant D2D links; $\mathcal{L} = \{1, 2, ..., L\}$.
- $\mathcal{C}$ is the set of cellular UEs; $\mathcal{C} = \{1, 2, ..., C\}$.
- $\mathcal{N}$ is the set of tenants; $\mathcal{N} = \{1, 2, ..., N\}$.
- $\mathcal{K}$ is the set of available resources; $\mathcal{K} = \{1, 2, ..., K\}$.

$\mathcal{C}$ contains all the cellular users that belong to different tenants and are consequently served by separate slices. It can be represented as follows: $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \ldots \mathcal{C}_N$, where $|\mathcal{C}_j| < C,\ \forall j \in \mathcal{C}$. Similarly, regarding the fused set of RBs $\mathcal{K}$, it consists of all tenants' radio resources, so it can be written as $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \ldots \mathcal{K}_N$, where $|\mathcal{K}_n| < K,\ \forall n \in \mathcal{N}$.

Further, we need to introduce the binary decision variable that indicates if a D2D link $l \in \mathcal{L}$ utilizes a specific RB k that belongs to one of the tenants' available resource pool. This can be mathematically represented as:

$$x_{lnk} = \begin{cases} 1, & \text{if D2D link } l \text{ uses RB } k \text{ of tenant } n \\ 0, & \text{otherwise.} \end{cases} \tag{5.11}$$

We further proceed with some important system model admissions to pave the way for the problem formulation. First, the path-loss is modeled as follows:

$$PL_{D2D} = 148 + 40 \log_{10} d \tag{5.12}$$

$$PL_{CUE} = 128.1 + 37.6 \log_{10} d \tag{5.13}$$

for D2D pairs and cellular users, respectively [94],[95]. Parameter d stands for the Euclidean distance and is expressed in kilometers. Additionally, the SINR at the D2D receiver of link l that uses RB $k = k_l$ of tenant n needs to be satisfied. If we denote by γ_{lnk_l} this value which translates for the receiver's need to correctly decode transmitted packets, this constraint can be practically expressed as follows:

$$\gamma_{lnk_l} = \frac{h_{ll}^{nk_l} P_l}{\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} x_{lnk} h_{cl}^{nk} P_c + Q + \sigma^2} \geq \gamma_{th} \tag{5.14}$$

where $h_{ll}^{nk_l}$ is the link gain (path-loss and slow fading dependent) of the l^{th} D2D pair, and P_l is the transmission power of the D2D transmitter over this RB. In the denominator, h_{cl}^{nk} expresses the link gain between the transmitting CUE c and the receiver of D2D link l when using RB k of tenant n (CUE-D2D interference will be developed when $k = k_l$ and $x_{lnk} = 1$). As it will be mentioned in the sequel, we consider that D2D links will be using orthogonal RBs among each other, i.e. the received interference of the D2D links will be only deriving from CUEs (and vice versa). Lastly, σ^2 denotes the lump sum power of background/thermal noise and Q the co-channel interference from other cells (if existent). In that case, we assume that inter-cell interference can be controlled via the application of powerful ICIC techniques, thus, we are focusing on a single-cell scenario ($Q = 0$) where the main part of interference (i.e. intra-cell) is effectively captured.

The SINR threshold ($\tilde{\gamma}_{th}$) needs to be also satisfied for the cellular transmissions that utilize the same RB (e.g. $k = k_c$) with a D2D link l during the uplink session. This constraint can be written as follows:

$$\frac{h_{cb}^{nk_c} P_c}{\sum_{l\in\mathcal{L}} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} x_{lnk} h_{lb}^{nk} P_l + Q + \sigma^2} \geq \tilde{\gamma}_{th} \tag{5.15}$$

where P_c is the transmission power of a cellular user, h_{cb}^{nk} is the link gain between the CUE c that belongs to tenant n and its associated BS b when using RB k, whereas h_{lb}^{nk} accounts for the link gain between the D2D transmitter of link l and the BS b that transmit/receive over the same channel k.

Considering the above definitions, the achievable rate for D2D link l that utilizes RB k of n tenant can be calculated according to the well-known Shannon capacity formula:

$$R_{lnk} = B_{RB} \log_2 \left(1 + \gamma_{lnk}\right) \tag{5.16}$$

where B_{RB} is the LTE-based resource block bandwidth (180 kHz) and γ_{lnk} is expressed in power ratio.

Lastly, even though we focus on the uplink scenario where communications happen according to the SC-FDMA principles, we herein consider that the RBs allocated per user can be non-adjacent ([107]) as the evolution of LTE towards 5G systems will eventually enable fully non-contiguous allocation. Considering this, we will practically provide an upper bound of the D2D-based rate performance. Following the previous admissions, the sum-rate maximization problem for cross-tenant D2D communications can be formulated as follows:

$$\underset{x}{max} \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_{lnk} x_{lnk} \tag{5.17}$$

subject to:

$$\sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} x_{lnk} h_{nk}^{lb} P_l \tilde{\gamma}_{th} \leq$$

$$- \left(\tilde{\gamma}_{th}(\sigma^2 + Q) - h_{nk_c}^{cb} P_c \right), \forall c \in \mathcal{C}, \; k_c \in \mathcal{K} \tag{5.17a}$$

$$\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_{lnk} x_{lnk} \geq R_l^{th}, \; \forall l \in \mathcal{L} \tag{5.17b}$$

$$\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} x_{lnk} \geq 1, \; \forall l \in \mathcal{L} \tag{5.17c}$$

$$\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} x_{lnk} \leq \Gamma, \; \forall l \in \mathcal{L} \tag{5.17d}$$

$$\sum_{l \in \mathcal{L}} x_{lnk} \leq 1, \; \forall n \in \mathcal{N}, \forall k \in \mathcal{K} \tag{5.17e}$$

$$x_{lnk} \in \{0, 1\}, \; \forall l \in \mathcal{L}, \; \forall n \in \mathcal{N}, \; \forall k \in \mathcal{K}. \tag{5.17f}$$

Constraint (5.17a) ensures that each cellular transmission's SINR doesn't fall below a predefined value $\tilde{\gamma}_{th}$, whereas (5.17b) guarantees the minimum rate requirement for each D2D link $l \in \mathcal{L}$. Constraints (5.17c) and (5.17d) account for the radio resource allocation of each D2D link; the former ensures that each D2D pair will be assigned with at least one RB to satisfy its transmission needs, whereas the latter upper bounds the resources used by each link to Γ to avoid any resource deficiency for some DUEs. Then, the restriction that each RB can be used by only one D2D link is realized by (5.17e). Finally, (5.17f) denotes the binary nature of the decision variable.

Finally, it is obvious that the overall rate achieved by a D2D link i is $r_i^{tot} = \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_{lnk} x_{lnk}$ and depends on the value assignment of the decision vector $\mathbf{x}$ that solves this optimization problem.

5.3.4 Heuristic Algorithm

Complementary to the previous technique, a heuristic algorithm is proposed to seek for a low-complexity, near optimal solution for D2D users that belong to different tenants. One of its chief characteristics is that it tries to achieve a fairly balanced, inter-slice resource allocation by sequentially running for D2D receivers that belong to different tenants. Its resource assignment rationale is based on allocating the resource blocks that provide the best channel conditions to each D2D link (in a sorted way) following the aforementioned sequential mode. Herein, it has to be noted that in order for some RB to be assigned to a D2D link, first, it must be an unallocated one (among D2Ds) and second, not to degrade the performance of the cellular uplink transmission that utilizes the same RB. Then, the algorithm iterates over and over, until one of the following conditions is violated: (i) all D2D users reach their upper RB usage limit (i.e. Γ used RBs), (ii) the fused RB pool is fully utilized by the active D2D transmissions, or (iii) DUEs' SINR requirements over the remaining RBs are not satisfied. The explained method is outlined in Algorithm 5.

Algorithm 5: INTER-TENANT HEURISTIC ALGORITHM

Data: CUEs'-DUEs' location coordinates, Γ.

$*$ **Assumptions** :

- Each $c \in \mathcal{C}$ is assumed to be served by a specific tenant $n \in \mathcal{N}$.
- Orthogonal, round-robin based resource allocation for each $c \in \mathcal{C}_1 \cup \mathcal{C}_2$ is applied.

$*$ **Inter-tenant D2D resource allocation steps** :

Step 1 : *D2D RB assignment*

```
repeat
    l = 1;   ▷ D2D link identifier.
    while l ≤ |L| do
        if (l mod 2 ≡ 1) then
            • slice_id = A;   ▷ slice/tenant identifier.
            • l_A = l;   ▷ D2D pair where the receiver is subscribed to tenant A.
            • Find the D2D-CUE combination that results in the maximum
            possible SINR for the D2D link;
            • Allocate CUE's assigned RBs to link l_A ∈ L;
            • Remove allocated RBs from available resource pool K;
            • l = l + 1;
        else
            • slice_id = B;
            • l_B = l;   ▷ D2D pair where the receiver is subscribed to tenant B.
            • Repeat the same procedure and update RB pool K;
            • l = l + 1;
        end
    end
until ⟨RB pool is fully used OR Γ RBs are assigned for all D2Ds OR no
more RBs can be assigned due to SINR requirements' violation⟩
```

* Note: (l mod 2) ensures the sequential RB allocation of D2D users for the example of two slices.

Step 2 : *Sum-rate estimation*

- $R_{tot} = \sum_{l \in \mathcal{L}} R_l$; $\triangleright$ R_l is the achieved rate $\forall l \in \mathcal{L}$.

5.3.5 Performance Evaluation

In this section, a set of evaluation results is provided to shed light on the performance of the proposed inter-tenant D2D sum-rate optimization problem compared to legacy approaches and heuristic solutions.

Compared methodology

In this subsection, a number of different D2D-based resource allocation techniques for inter-tenant communications are briefly described. These techniques constitute the compared methodology through which the results to follow are produced.

1. **Inter-tenant optimal**: The proposed method was detailed in subsection 5.3.3. As previously explained, it yields optimal sum-rate performance for inter-tenant D2D users via a powerful ILP solution that virtually fuses the provided to the tenants RB pools and orchestrates the links' resource assignment.

2. **Inter-tenant heuristic**: Complementary to the previous technique, a heuristic algorithm is proposed to seek for a low-complexity, near optimal solution for D2D users that belong to different tenants. One of its chief characteristics is that it tries to achieve a fairly balanced, inter-slice resource allocation by sequentially running for D2D receivers that belong to different tenants. Its resource assignment rationale is based on allocating the resource blocks that provide the best channel conditions to each D2D link (in a sorted way) following the aforementioned sequential mode. Herein, it has to be noted that in order for some RB to be assigned to a D2D link, first, it must be an unallocated one (among D2Ds) and second, not to degrade the performance of the cellular uplink transmission that utilizes the same RB. Then, the algorithm iterates over and over, until one

of the following conditions is violated: (i) all D2D users reach their upper RB usage limit (i.e. Γ used RBs), (ii) the fused RB pool is fully utilized by the active D2D transmissions, or (iii) DUEs' SINR requirements over the remaining RBs are not satisfied. The explained method is outlined in Algorithm 5.

3. **Intra-tenant optimal**: With this technique, problem (5.17) is decoupled into two separate resource allocation problems for the two different tenants. This means that each tenant solves separately the sum-rate optimization problem for its subscribed D2D users that initiate direct peer communications, based on its corresponding dedicated slice resources. Due to the restricted RB availability for the different tenants, this method, even though it is able to provide optimal sum-rate performance from each tenant's side, it is expected to provide a sub-optimal solution in overall.

4. **Intra-tenant heuristic**: Depending on the number of DUEs that are subscribed to a specific tenant and initiate a number of inter-slice connections, this tenant is the one to provide the corresponding direct communications with the suitable RB pool to satisfy their transmission needs. To this end, each one of the tenants allocates resources to the corresponding D2Ds in a greedy and sorted manner according to best-given channel conditions. This method is similar to the inter-tenant heuristic approach but again is decoupled as it needs to be solved by each different tenant for the subscribed users.

Simulation setup

The considered system was modeled in MATLAB, following the LTE-A milestones and corresponding network parameters and standards. All the produced results derived after averaging over 1000 Monte Carlo simulations which have

Table 5.2: Simulation Parameters

Parameter	**Value**
Cell layout	Hexagonal grid
Number of tenants ($\mathcal{N}$)	2
CUEs-D2Ds distribution	Uniform
Macro cell radius	400 m
Maximum D2D link range	100 m
Number of CUEs (C)	50
Number of D2D links (L)	[10,40]
D2D Path-Loss model	$148 + 40 \log_{10} d$
CUE-BS Path-Loss model	$128.1 + 37.6 \log_{10} d$
Maximum CUEs' power	20 dBm
Maximum DUEs' power	15 dBm
Maximum number of RBs (Γ)	4
Shadowing standard deviation	8 dB
Noise power spectral density	-174 dBm/Hz
System bandwidth (BW)	10 MHz

been executed on a *Intel(R) Core(TM) i7-6500 at 2.50 GHZ and 8 GB RAM* machine.

Regarding the topology, it consists of a hexagonal single cell with randomly distributed cellular users and D2D links. Each D2D link consists of two users that are assumed to belong to different tenants. Also, the number of cellular users being subscribed to different tenants is varying and considers the tenants' disparities in terms of number of subscriptions. Without loss of generality, two tenants, A and B, with separate slices are considered. All simulation parameters are listed in Table 5.2.

Results

Due to the load discrepancies and divergent number of subscriptions that might characterize the two or more tenants (either MVNOs or MNOs), the slice that each one is assigned is expected to be different (i.e., having heterogeneous slices). Popular tenants can be normally assigned with more resources to serve the high number of subscribed users compared to less popular ones. To this end, we consider the case where a popular tenant (denoted as tenant A hereafter) is allocated with double-sized RB pool to serve its users. Initially, we assume that for both tenants all the radio resources are occupied by a number of cellular UEs according to a Round Robin scheduling. Figure 5.10 depicts the sum-rate performance for inter-tenant (inter-slice) D2D links in relation to varying number of them. On average, almost 11.3% sum-rate gain is achieved by making use of the fused RB pools of the two tenants (inter-tenant optimal) compared to the case that a D2D link can be assigned resources only from the resource pool that belongs to the slice to which the user that initiates the direct communication is subscribed (intra-tenant optimal). The maximum performance gap among the illustrated scenarios is met in the case of 16 D2D links, where 12.5% higher sum-rate is achieved with the inter-tenant optimization solution. Further, compared to the heuristic inter-tenant approach, the optimal solution is averagely 8.45% better and gradually behaves better with the increase of inter-tenant D2D links. Also, the intra-tenant heuristic algorithm falls short compared to the above-mentioned approaches and it exhibits a maximum of more than 18% sum-rate degradation in comparison to the optimal solution. Last, for all the considered approaches, the sum-rate performance drop that is observed in the two last cases is explained by the increase of interference to/from cellular users as the resource availability gets more restricted.

Considering the same case study, the cumulative distribution function (CDF) of the achieved SINR values for the inter-tenant D2D links is represented in figure 5.11. Indicatively, in the 50^{th} percentile, the inter-tenant optimal solution's SINR for D2Ds is 29.1 dBs, whereas the corresponding values for the inter-tenant heuristic, the intra-tenant optimization and the intra-tenant heuristic are 26, 25.7

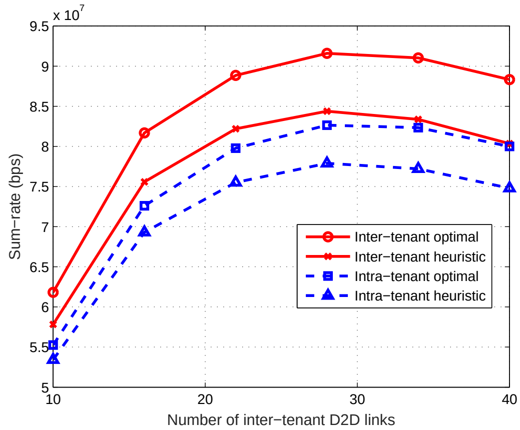


Figure 5.10: Sum-rate comparison for inter-slice communications in relation to varying number of D2D links.

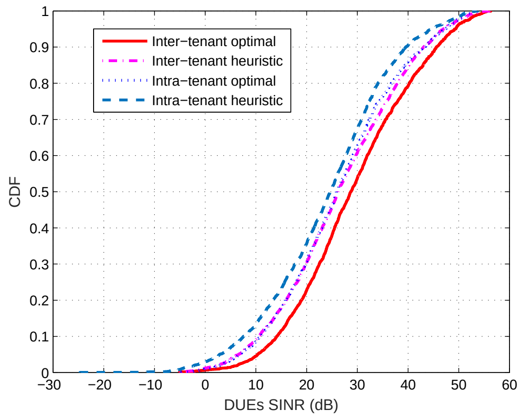


Figure 5.11: SINR-based CDF for cross-tenant D2D links.

and 23.8 dBs, respectively. This can be interpreted as more than 2 times higher SINR power ratio compared to the inter-tenant heuristic, 2.18 better compared to the intra-tenant optimal. Last, the intra-tenant heuristic estimated SINR is almost 3.39 times worse compared to the optimal value.

Additionally, we consider the scenario where the two tenants are characterized by the same RB availability but with different utilization levels (active cellular transmissions per case). To this direction, tenant's A radio resources are supposed to be fully occupied by its subscribed CUEs, whereas tenant's B resource block availability ranges from 20% to 100%. Figure 5.12 depicts the sum-rate performance of all compared methods in relation to the normalized resource uti-

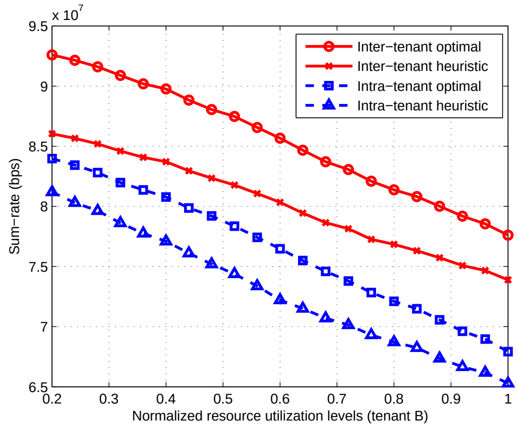


Figure 5.12: Sum-rate performance in relation to RB utilization levels. The total number of D2D inter-tenant links is fixed to 20.

lization of tenant B. As expected, while the resource occupancy increases, the sum-rate decreases for all methods as new interference patterns between cellular and D2D users arise. However, the performance gap between the inter-tenant optimal and heuristic techniques lessens with the increase of the RB occupation levels for tenant B, as opposed to the rest of the methods where the gap slightly increases. Quantifying the above observations, the inter-tenant proposed optimization formulation outperforms the heuristic, the intra-tenant optimal and heuristic algorithms in an average of almost 6.5%, 12% and 17.5%, respectively. When both tenants' RB pools are fully utilized (reaching 100% of resource utilization), all D2D links are reusing part of the cellular spectrum that CUEs occupy. In that case, the inter-tenant optimal solution achieves its peak sum-rate gain compared to the intra-tenant methodology; a 14.3% improvement is observed over the intra-tenant optimal and 18.7% over the related heuristic, respectively. This result can be deemed as highly interesting because the maximum gains take place when needed, i.e. during network congestion episodes.

The sum-rate performance of D2D users in relation to maximum link length is further evaluated. By increasing the allowable limit of maximum D2D link length, the sum-rate performance of all compared methods follows a decreasing trend. This is expected, as the increase of D2D link length implies higher SINR

Table 5.3: Algorithm running times

Algorithm \ Number of links	I = 10	I = 20	I = 30	I = 40
Inter-tenant optimal	0.4516 s	0.0465 s	0.0392 s	0.0342 s
Inter-tenant heuristic	0.0342 s	0.0182 s	0.0086 s	0.0048 s
Intra-tenant optimal	0.0786 s	0.0412 s	0.0351 s	0.0332 s
Intra-tenant heuristic	0.0282 s	0.0204 s	0.0169 s	0.0055 s

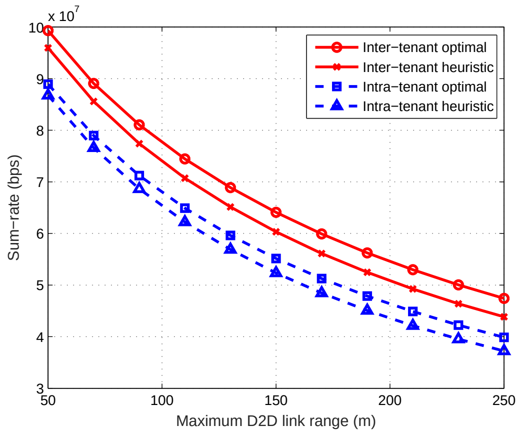


Figure 5.13: Sum-rate performance in relation to the maximum allowable D2D link range.

degradation not only due to path-loss and shadowing effects but also due to different emerging interference patterns. The performance gains become more clear for the largest values of link lengths. Notably, the inter-tenant optimization problem is almost 8.15%, 18.9% and 27.4% better in terms of sum-rate when compared to inter-tenant heuristic, intra-tenant optimal and heuristic solutions, respectively. Also, on aggregate, the inter-tenant heuristic method is the one that provides again the closest among all performance as it falls short almost 6% in terms of throughput compared to the optimal one. Considering the rest, the optimal solution provides an average gain of almost 16% and more than 22% compared to the intra-tenant optimal and heuristic techniques.

Finally, in order to give a glimpse of their computational complexity, the running times of the aforementioned algorithms are listed in Table 2. It is shown that with the increase of the number of D2D inter-tenant links, the running times of all compared algorithms decrease. This can be explained by the fact that when less D2D links exist in the topology, the probability that all or many of them will utilize the maximum assignable number of resources (Γ) to increase as much as possible their rate performance raises. Thus, the number of combinations for the orthogonal assignment of D2D links increases. Although the inter-tenant optimal solution is proven to be the most complex, its running time remains in acceptable levels.

5.4 Summary

First, in this chapter, a resource slicing framework for radio access networks that take explicitly into account D2D communication pairs is proposed. In specific, a linear integer mathematical solution that provides upper bounds on the achievable performance as well as low-complexity but sub-optimal algorithms that are amenable to real-time implementations are proposed. Via an extensive set of numerical investigations we showed that the applied framework has the potential of significantly improving the throughput performance of D2D links whilst causing the least interference to the receiving BS compared to previously proposed RAN resource slicing techniques.

In the second part of the chapter, the problem where the two UEs of a D2D communication belong to different service and/or network providers is investigated. Under the assumption of full virtualized core/access networks, an optimization framework for efficiently use of virtualized resources across different tenants enabled by a cross-tenant controller is provided. Via a wide set of numerical investigations it has been shown that significant throughput gains of over 10% compared to legacy solutions can be achieved for inter-tenant D2D communications. These results also reinforce the need for implementing a cross-slice coordinator, which can be considered as an extension to SDN/NFV frameworks, in order to efficiently utilize the scarce wireless resources across separate network slices.

Chapter 6

Conclusions and Future Work

6.1 Concluding Remarks

The evolution towards 5G is accompanied with the ongoing integration of disruptive technologies that are not only able to improve current networks' performance but also introduce new business opportunities across several industries. Motivated by the high achievable data rates coming from local area services, such as file distribution, D2D communication has emerged as the prevalent enabler of proximity-based peer-to-peer connections in future wireless networks. By allocating the cellular spectrum for both cellular and D2D communications, spectral efficiency and network capacity can be immensely enhanced as well as the traffic developed on the base station side can be significantly eased. However, the scarce resource availability as well as the developed interference between D2D and cellular links might be a hindering factor towards fully exploiting the merits offered by D2D. To this end, efficient cell association and radio resource allocation techniques have to be devised in order to optimize the coexistence of D2D and cellular users as well as disruptively improve their performance.

Compared to the existing norm, cell association has to be redefined in order to further consider the nature of D2D communications. There might be a case where two users that constitute a D2D pair associate with two different base stations. In such a scenario, the two base stations need to exchange information in order to coordinate the D2D link establishment, operation and resource allocation. However, this base stations' intercommunication leads to increased signalling overhead, delay and latency that could be eliminated if the two communicating users were associated with only one base station. Based on this principal assumption, Chapter 2 splits the problem of cell association for D2D communications in two parts: first, in the overlay case (dedicated licensed spectrum for D2D communication), and second, in the underlay one (cellular and D2D users share the whole licensed spectrum). In the former, the D2D cell association problem based on the concept of uplink and downlink decoupling is investigated to highlight the supremacy of the decoupled association compared to other D2D-aware cell association techniques in terms of interference and power consumption. In the latter, a cell association optimization framework for D2D communications, namely *MOCA*, is applied to account for increased network capacity with respect to radio resource constraints, interference as well as network traffic. This framework constitutes the basis of an alternative, near optimal solution to the NP-hard problem of joint cell association and radio resource allocation in a D2D underlaying cellular networks' scenario. Then, an iterative randomized heuristic algorithm, namely *i-RRA*, is proposed to not only highlight the contributing points and effectiveness of the cell association scheme but also provide a fast and efficient solution in terms of sum-rate performance compared to baseline techniques, as proven via extensive numerical investigations.

In chapter 3, we proceed with investigating resource allocation potentials in another special D2D-based networking scenario. In specific, the integration of relay-aided D2D communications as an underlay in cellular networks is examined. Assuming that all user devices are relay-enabled and can operate as intermediate nodes to assist in providing reliable transmissions, we proposed a resource allocation method based on the notion of genetic algorithms to maximize the network throughput performance with respect to interference. Genetic algorithms, as part of bio-inspired meta-heuristics are proven to provide low-complexity, competitive solutions in the telecommunications' industry among all. The simulation results showed that via genetic algorithmic solutions the aggregate network throughput can be improved and the interference coming from cellular connections can be controlled when compared to other baseline resource allocation techniques.

Chapter 4 elaborates on a futuristic aspect that is expected to constitute significant part of the emerging 5G networks, that of inter-operator communications. To be more specific, we focused on a scenario where the two users of a D2D pair belong to two different operators and it is not clear how it will be allocated with radio resources to satisfy its communication needs. Under the assumption of fully virtualized core/access networks due to the gains offered by the virtualization of wireless radio resources, where multiple tenants are envisaged to share the substrate physical infrastructure, we proposed a sum-rate optimization solution for D2D communications that effectively utilizes the virtualized resources across different tenants based on a cross-tenant, centralized controller. This proposal does not only outperform legacy approaches in terms of throughput performance (over 10% sum-rate gain can be achieved) but also introduces the novel concept of cross-tenant coordinator which can be considered as an additive component in SDN/NFV frameworks to disruptively utilize the scarce radio resources.

Finally, Chapter 5 studies the problem of content caching and file dissemination via direct D2D links in cellular networks based on the notion of network coding. By considering cache-enabled user terminals, we proposed a network coded, cooperative cache management method that considers the compression of the existing files in a user's cache to store more contents of interest. This technique is then compared with a number of popular cache orchestration methods where its performance in terms of traffic offloading from the base station is illustrated. Numerical results showed that not only most of the content requests are being satisfied by local D2D communications in the long run but also the compression-based technique can effectively alleviate the BS traffic in a shorter time window.

6.2 Future Work

6.2.1 Mobility-enabled D2D communications

Vehicle-to-Vehicle (V2V) communication is one of the most attracting D2D-related applications as a form of mobility enabled direct communications. It is expected to play a significant role in 5G networks as a major vertical that will eventually pave the way for creating new revenue streams for network operators and vendors as well as introduce novel business opportunities. While this communication type was so far used for exchanging mainly safety-related messages with the infrastructure (V2X), future vehicular communications are envisaged to enable a multitude of advanced features, such as automated driving, augmented reality functions and advanced infotainment where car drivers/passengers would be allowed to stream high definition and bulky videos in the car. The aforementioned features imply varying traffic requirements; on the one side, reliability and low latency demanding traffic (e.g. traffic signal alerts, beacons etc.), and, on the other side, high data rate demanding traffic (e.g. multimedia). In order to allow for such demanding services and applications, enhancements to the existing network architecture and further design considerations need to be considered. To this end, innovative radio resource management techniques that take into account the different traffic demands as well as the resource availability constraints need to be devised. Also, the aspect of interoperability, and specifically how multiple terminals that are subscribed to different operators will communicate between each other and radio resources will be orchestrated is of critical importance. Finally, caching will be another focal point in mobility-oriented scenarios towards optimizing the placement of popular content as closer as possible to the user end, aiming at reducing the critical information exchange times.

6.2.2 Virtualized Resource Sharing in Cloud-RAN based 5G Networks

Cloud radio access network (C-RAN) is a prominent technology that aims at improving spectrum efficiency by managing a large number of cell sites in a centralized and cooperative fashion as it aggregates all deployed base stations' computational resources into a central pool. Briefly, this real-time virtualization architecture collects the radio frequency signals from a number of distributed antennas via the deployed remote radio heads (RRHs) and transmits them to the baseband unit (BBU) pool through optical transport network (OTN) in order to achieve cooperative multi-point processing. The integration of D2D communications in C-RAN based networks is quite challenging since, other than spectrum utilization, it can ease the congestion on the fronthaul in times of high traffic episodes. To this direction, the focus has to be turned on exploring the potential gains of sharing virtualized resources between multiple tenants compared to the traditional virtualization view where resource isolation, i.e. assignment of orthogonal sets of radio resources for a certain period to each tenant and with respect to specific services' requirements, should be also conserved.

References

[1] CISCO, "Global Mobile Data Traffic Forecast Update, 2015—2020," in *Cisco White Paper*, February 2016.

[2] Y. Meng, C. Jiang, H. H. Chen, and Y. Ren, "Cooperative device-to-device communications: Social networking perspectives," *IEEE Network*, vol. 31, no. 3, pp. 38–44, May 2017.

[3] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5g," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, May 2016.

[4] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A Survey on 3GPP Heterogeneous Networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, June 2011.

[5] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5G on the Horizon: Key Challenges for the Radio-Access Network," *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 47–53, September 2013.

[6] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5g multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 118–127, June 2014.

[7] U. N. Kar and D. K. Sanyal, "An overview of device-to-device communication in cellular networks," *ICT Express*, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405959517301467

[8] 3GPP TR 36.843, "Technical Specification Group Radio Access Network; Study on LTE device to device proximity services," Tech. Rep., March 2014.

[9] 3GPP, "UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE)," Tech. Rep., October 2006.

[10] ——, "Proposal for Candidate Radio Interface Technologies for IMT-Advanced Based on LTE Release 10 and Beyond (LTE-Advanced)," Tech. Rep., October 2009.

[11] S. Mumtaz and J. Rodriguez, *Smart Device to Smart Device Communication.* Springer Publishing Company, Incorporated, 2014.

[12] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, December 2009.

[13] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, and Z. Turanyi, "Design Aspects of Network Assisted Device-to-Device Communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, March 2012.

[14] T. Ihalainen, P. Janis, Z. Li, M. Moisio, V. Nurmela, M. Uusitalo, C. Wijting, and O. Yilmaz, "Flexible Scalable Solutions for Dense Small Cell Networks," in *Wireless World Research Forum (WWRF)*, April 2013.

[15] P. Herath, W. A. Krzymien, and C. Tellambura, "A Novel Base Stations-Mobile Stations Association Policy for Cellular Networks," in *IEEE 80th Vehicular Technology Conference (VTC Fall)*, September 2014, pp. 1–5.

[16] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell Association and Interference Coordination in Heterogeneous LTE-A Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1479–1489, December 2010.

[17] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[18] 3GPP TSG RAN WG1 Meeting-62 36.843, "Potential Performance of Range Expansion in Macro-Pico Depoloyment (R1-104355)," Tech. Rep., August 2010.

[19] I. Guvenc, "Capacity and Fairness Analysis of Heterogeneous Networks with Range Expansion and Interference Coordination," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1084–1087, October 2011.

[20] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Outage Probability for Heterogeneous Cellular Networks with Biased Cell Association," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, December 2011, pp. 1–5.

[21] ——, "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3484–3495, October 2012.

[22] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, "User Association in 5G Networks: A Survey and an Outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1018 – 1044, January 2016.

[23] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and Uplink Decoupling: A Disruptive Architectural Design for 5G networks," in *IEEE Global Communications Conference (GLOBECOM)*, December 2014, pp. 1798–1803.

[24] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to Decouple the Uplink and Downlink in Cellular Networks and How To Do it," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 110–117, March 2016.

[25] J. Liu, Y. Kawamoto, H. Nishiyama, N. Kato, and N. Kadowaki, "Device-to-Device Communications Achieve Efficient Load Balancing in LTE-Advanced Networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 57–65, April 2014.

[26] S. Xiao, D. Feng, Y. Yuan-Wu, G. Y. Li, W. Guo, and S. Li, "Optimal Mobile Association in Device-to-Device-Enabled Heterogeneous Networks," in *IEEE 82nd Vehicular Technology Conference (VTC Fall)*, September 2015, pp. 1–5.

[27] J. Wang, D. Zhu, C. Zhao, J. C. Li, and M. Lei, "Resource Sharing of Underlaying Device-to-Device and Uplink Cellular Communications," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1148 – 1151, 2013.

[28] P. Phunchongharn, E. Hossain, and D. Kim, "Resource Allocation for Device-to-Device Communications Underlaying LTE-Advanced Networks," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 91–100, August 2013.

[29] P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl, "Interference-Aware Resource Allocation for Device-to-Device Radio Underlaying Cellular Networks," in *IEEE 69th Vehicular Technology Conference (VTC)*, April 2009, pp. 1–5.

[30] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient Resource Allocation for Device-to-Device Communication Underlaying LTE Network," in *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, October 2010, pp. 368–375.

[31] T. Chen, G. Charbit, and S. Hakola, "Time Hopping for Device-To-Device Communication in LTE Cellular System," in *IEEE Wireless Communication and Networking Conference (WCNC)*, April 2010, pp. 1–6.

[32] H. Wang and X. Chu, "Distance-constrained Resource-Sharing Criteria for Device-to-Device Communications Underlaying Cellular Networks," *Electronics Letters*, vol. 48, no. 9, pp. 528–530, April 2012.

[33] B. Wang, L. Chen, X. Chen, X. Zhang, and D. Yang, "Resource Allocation Optimization for Device-to-Device Communication Underlaying Cellular Networks," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–6.

[34] X. Zhu, S. Wen, G. Cao, X. Zhang, and D. Yang, "QoS-based Resource Allocation Scheme for Device-to-Device (D2D) Radio Underlaying Cellular Networks," in *19th International Conference on Telecommunications (ICT)*, April 2012, pp. 1–6.

[35] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, and B. Jiao, "Interference-Aware Resource Allocation for Device-to-Device Communications as an Underlay Using Sequential Second Price Auction," in *IEEE International Conference on Communications (ICC)*, June 2012, pp. 445–449.

[36] C. Xu, L. Song, Z. Han, D. Li, and B. Jiao, "Resource Allocation Using a Reverse Iterative Combinatorial Auction for Device-to-Device Underlay Cellular Networks," in *2012 IEEE Global Communications Conference (GLOBECOM)*, December 2012, pp. 4542–4547.

[37] F. Wang, C. Xu, L. Song, Q. Zhao, X. Wang, and Z. Han, "Energy-Aware Resource Allocation for Device-to-Device Underlay Communication," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 6076–6080.

[38] D. Wu, J. Wang, R. Q. Hu, Y. Cai, and L. Zhou, "Energy-Efficient Resource Sharing for Mobile Device-to-Device Multimedia Communications," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2093–2103, June 2014.

[39] C.-H. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen, "Resource Sharing Optimization for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2752–2763, August 2011.

[40] T. Han, R. Yin, Y. Xu, and G. Yu, "Uplink Channel Reusing Selection Optimization for Device-to-Device Communication Underlaying Cellular Networks," in *IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, September 2012, pp. 559–564.

[41] Z. Liu, T. Peng, H. Chen, and W. Wang, "Optimal D2D User Allocation Over Multi-Bands Under Heterogeneous Networks," in *2012 IEEE Global Communications Conference (GLOBECOM)*, December 2012, pp. 1339–1344.

[42] J. Han, Q. Cui, C. Yang, and X. Tao, "Bipartite Matching Approach to optimal Resource Allocation in Device to Device Underlaying Cellular Network," *Electronics Letters*, vol. 50, no. 3, pp. 212–214, January 2014.

[43] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, "Relay-based Deployment Con-

cepts for Wireless and Mobile Broadband Radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, September 2004.

[44] Y. Pei and Y. C. Liang, "Resource Allocation for Device-to-Device Communications Overlaying Two-Way Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3611–3621, July 2013.

[45] B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum Sharing Scheme Between Cellular Users and Ad-hoc Device-to-Device Users," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1038–1049, March 2013.

[46] M. Hasan, E. Hossain, and D. I. Kim, "Resource Allocation Under Channel Uncertainties for Relay-Aided Device-to-Device Communication Underlaying LTE-A Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2322–2338, April 2014.

[47] M. Hasan and E. Hossain, "Distributed Resource Allocation for Relay-Aided Device-to-Device Communication: A Message Passing Approach," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6326–6341, November 2014.

[48] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An Overview of 3GPP Device-to-Device Proximity Services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, April 2014.

[49] C. Liu, P. Whiting, and S. Hanly, "Joint Resource Allocation and User Association in Downlink Three-Tier Heterogeneous Networks," in *IEEE Global Communications Conference*, December 2014, pp. 4232–4238.

[50] C. Vlachos and V. Friderikos, "Optimal Device-to-Device Cell Association and Load Balancing," in *IEEE International Conference on Communications (ICC)*, London, UK, June 2015.

[51] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[52] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (EUTRA) Physical Layer Procedures (Release 10), v10.0.1," Tech. Rep., December 2010.

[53] O. Yilmaz, Z. Li, K. Valkealahti, M. Uusitalo, M. Moisio, P. Lunden, and C. Wijting, "Smart Mobility Management for D2D Communications in 5G Networks," in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2014, pp. 219–223.

[54] M. Coupechoux and J. M. Kelif, "How To Set the Fractional Power Control Compensation Factor in LTE?" in *34th IEEE Sarnoff Symposium*, May 2011, pp. 1–5.

[55] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator Controlled Device-to-Device Communications in LTE-Advanced Networks," *IEEE Wireless Communications*, vol. 19, no. 3, pp. 96–104, June 2012.

[56] 3GPP TR 22.803, "Feasibility Study for Proximity Services (ProSe)," Tech. Rep., June 2013.

[57] G. Wood and J. Robertson, "Buffon Got It Straight," *Statistics & Probability Letters*, vol. 37, no. 4, pp. 415 – 421, March 1998.

[58] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, November 2011.

[59] L. Melki, S. Najeh, and H. Besbes, "Radio Resource Allocation Scheme for Intra-Inter-Cell D2D Communications in LTE-A," in *IEEE Interna-*

tional Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), August 2015, pp. 1515–1519.

[60] J. Huang, Y. Yin, Y. Zhao, Q. Duan, W. Wang, and S. Yu, "A Game-Theoretic Resource Allocation Approach for Intercell Device-to-Device Communications in Cellular Networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 4, pp. 475–486, October 2016.

[61] S. Zhang, J. Liu, N. Kato, H. Ujikawa, and K. Suzuki, "Average Rate Analysis for a D2D Overlaying Two-Tier Downlink Cellular Network," in *IEEE ICC*, June 2015, pp. 3376–3381.

[62] J. Liu, N. Kato, H. Ujikawa, and K. Suzuki, "Device-to-Device Communication for Mobile Multimedia in Emerging 5G Networks," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMMCAP)*, vol. 12, no. 5s, pp. 75:1–75:20, September 2016.

[63] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System Architecture and Key Technologies for 5G Heterogeneous Cloud Radio Access Networks," *IEEE Network*, vol. 29, no. 2, pp. 6–14, March 2015.

[64] H. S. Chae, J. Gu, B.-G. Choi, and M. Chung, "Radio Resource Allocation Scheme for Device-to-Device Communication in Cellular Networks Using Fractional Frequency Reuse," in *17th Asia-Pacific Conference on Communications (APCC)*, October 2011, pp. 58–62.

[65] T. Novlan, J. Andrews, I. Sohn, R. Ganti, and A. Ghosh, "Comparison of Fractional Frequency Reuse Approaches in the OFDMA Cellular Downlink," in *IEEE GLOBECOM*, December 2010, pp. 1–5.

[66] J. Branke, K. Deb, K. Miettinen, and R. Slowinski, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, 2008, vol. 5252.

[67] R. Marler and J. Arora, "The Weighted Sum Method for Multi-objective Optimization: New Insights," *Structural and Multidisciplinary Optimization*, vol. 41, no. 6, pp. 853–862, 2010.

[68] I. Foster, *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering.* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1995.

[69] G. Fodor, D. D. Penda, M. Belleschi, M. Johansson, and A. Abrardo, "A Comparative Study of Power Control Approaches for Device-to-Device Communications," in *IEEE ICC*, June 2013, pp. 6008–6013.

[70] D. Zhao and K.-W. Chin, "Approximation Algorithms for Interference Aware Broadcast in Wireless Networks," in *14th IEEE International Symposium and Workshops of World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2013, pp. 1–9.

[71] J. Johnson and Y. Rahmat-Samii, "Genetic Algorithm Optimization of Wireless Communication Networks," *IEEE Antennas and Propagation Society International Symposium*, June 1995.

[72] Y. Zhenhua, Y. Guangwen, L. Shanwei, and Z. Qishan, "A Modified Immune Genetic Algorithm for Channel Assignment Problems in Cellular Radio Networks," in *Int. Conference on Intelligent System Design and Engineering Application (ISDEA)*, vol. 2, Oct 2010, pp. 823–826.

[73] M. Lima, A. Araujo, and A. Cesar, "Adaptive Genetic Algorithms for Dynamic Channel Assignment in Mobile Cellular Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2685–2696, Sept 2007.

[74] T. Fang and L. Chau, "GOP-based Channel Rate Allocation Using Genetic Algorithm for Scalable Video Streaming over Error-prone Networks," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1323–1330, June 2006.

[75] S. Lembo, A. Dowhuszko, and O. Tirkkonen, "Joint Resource Allocation in Mobile Networks with Macro Cellular and Device-to-Device Communication," in *IEEE ICC Workshop*, June 2015, pp. 585–590.

[76] L. Su, Y. Ji, P. Wang, and F. Liu, "Resource Allocation Using Particle Swarm Optimization for D2D Communication Underlay of Cellular Networks," in *IEEE WCNC*, April 2013, pp. 129–133.

[77] S. Maghsudi and S. Stanczak, "Hybrid Centralized-Distributed Resource Allocation for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Transactions on Vehicular Technology*, no. 99, 2015.

[78] T. Riihonen, S. Werner, and R. Wichman, "Comparison of Full-Duplex and Half-Duplex Modes with a Fixed Amplify-and-Forward Relay," in *IEEE WCNC*, April 2009, pp. 13–17.

[79] L. Badia, A. Botta, and L. Lenzini, "A Genetic Approach to Joint Routing and Link Scheduling for Wireless Mesh Networks," *Ad Hoc Networks*, vol. 7, no. 4, pp. 654 – 664, 2009.

[80] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio Access Network Virtualization for Future Mobile Carrier Networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 27–35, July 2013.

[81] 3GPP TS 23.251, "Network Sharing; Architecture and Functional Description," Tech. Rep., 2012.

[82] Open Networking Foundation (ONF), "Software Defined Networking: The New Norm for Networks," in *ONF White Paper*, April 2012.

[83] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design Considerations for a 5G Network Architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, November 2014.

[84] F. Hu, Q. Hao, and K. Bao, "A Survey on Software-Defined Network and OpenFlow: From Concept to Implementation," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 2181–2206, Fourthquarter 2014.

[85] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network Function Virtualization: Challenges and Opportunities for Innovations," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 90–97, February 2015.

[86] R. Mijumbi, J. Serrat, J. L. Gorricho, N. Bouten, F. D. Turck, and R. Boutaba, "Network Function Virtualization: State-of-the-Art and Research Challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 236–262, Firstquarter 2016.

[87] A. Asadi, Q. Wang, and V. Mancuso, "A Survey on Device-to-Device Communication in Cellular Networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[88] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference-Aware Graph Based Resource Sharing for Device-to-Device Communications Underlaying Cellular Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013.

[89] M. Belleschi, G. Fodor, and A. Abrardo, "Performance Analysis of a Distributed Resource Allocation Scheme for D2D Communications," in *IEEE Global Communications Conference Workshops (GLOBECOM Wkshps)*, December 2011, pp. 358–362.

[90] Y. Cai, F. Yu, and C. Liang, "Resource Sharing for Software Defined D2D Communications in Virtual Wireless Networks with Imperfect NSI," in *IEEE Global Communications Conference (GLOBECOM)*, December 2014.

[91] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and Flexible Cellular Core Network Architecture," in *Proceedings of the Ninth ACM Conference on Emerging Network Experiments and Technologies*, December 2013.

[92] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software Defined Radio Access Network," in *Proc. of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, Aug. 2013.

[93] J. Qadir, N. Ahmed, and N. Ahad, "Building Programmable Wireless Networks: an Architectural Survey," *EURASIP Journal on Wireless Communications and Networking*, Oct. 2014.

[94] 3GPP TR 36.814, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," Tech. Rep., 2010.

[95] H. Pang, P. Wang, X. Wang, F. Liu, and N. Ngoc Van, "Joint Mode Selection and Resource Allocation Using Evolutionary Algorithm for Device-to-Device Communication Underlaying Cellular Networks," *Journal of Communications (JCM)*, vol. 8, no. 11, pp. 751–757, Nov. 2013.

[96] J. Gu, H. Yoon, J. Lee, S. Bae, and M. Chung, "A Resource Allocation Scheme for Device-to-Device Communications Using LTE-A Uplink Resources," *Pervasive and Mobile Computing*, vol. 18, pp. 104 – 117, Feb. 2015.

[97] M. Kalil, A. Shami, and Y. Ye, "Wireless Resources Virtualization in LTE Systems," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 363–368.

[98] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, Firstquarter 2015.

[99] V. Phan, L. Yu, K. Horneman, and S. Hakola, "D2D Communications Considering Different Network Operators," Octorber 2011, uS Patent App. 12/763,454. [Online]. Available: http://www.google.com/patents/US20110258327

[100] B. Cho, K. Koufos, R. Jäntti, Z. Li, and M. A. Uusitalo, "Spectrum Allocation for Multi-Operator Device-to-Device Communication," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 5454–5459.

[101] C. Vlachos and V. Friderikos, "Optimal Virtualized Resource Slicing for Device-to-Device Communications," in *IEEE Global Communications Conference (GLOBECOM)*, December 2015, pp. 1–7.

[102] P. Rost, I. Berberana, A. Maeder, H. Paul, V. Suryaprakash, M. Valenti, D. Wübben, A. Dekorsy, and G. Fettweis, "Benefits and Challenges of Virtualization in 5G Radio Access Networks," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 75–82, December 2015.

[103] 3GPP TS 22.951, "Technical Specification Group Services and System Aspects; Service aspects and requirements for network sharing (Release 13) ," Tech. Rep., December 2015.

[104] 3GPP TR 22.852, "Technical Specification Group Services and System Aspects; Study on Radio Access Network (RAN) sharing enhancements (Release 13)," Tech. Rep., September 2014.

[105] A. Moubayed, A. Shami, and H. Lutfiyya, "Wireless Resource Virtualization with Device-to-Device Communication Underlaying LTE Network," *IEEE Transactions on Broadcasting*, vol. 61, no. 4, pp. 734–740, December 2015.

[106] B. Wang, L. Chen, X. Chen, X. Zhang, and D. Yang, "Resource Allocation Optimization for Device-to-Device Communication Underlaying Cellular Networks," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–6.

[107] H. Wang, C. Rosa, and K. I. Pedersen, "Radio resource management for uplink carrier aggregation in lte-advanced," *EURASIP Journal on Wireless Communications and Networking*, no. 1, p. 121, 2015.

Appendix A

Related Publications

A.1 Journal papers

1. C. Vlachos and V. Friderikos, "MOCA: Multi-Objective Cell Association for Device-to-Device Communications", in *IEEE Transactions on Vehicular Technology (TVT)*, April 2017.

2. C. Vlachos, V. Friderikos and M. Dohler, "Optimal Virtualized Inter-Tenant Resource Sharing for Device-to-Device Communications in 5G Networks", in *ACM/Springer Mobile Networks & Applications (MONET)*, Special issue on: *Device-to-Device Communication in 5G Networks*, February 2017.

A.2 Conference papers

1. C. Vlachos and V. Friderikos, "Robust Randomized Resource Allocation for Device-to-Device Communication" in *IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, December 2014.

2. C. Vlachos and V. Friderikos, "Optimal Device-to-Device Cell Association and Load Balancing", in *IEEE International Conference on Communications (ICC)*, June 2015.

3. C. Vlachos and V. Friderikos, "Optimal Virtualized Resource Slicing for Device-to-Device Communications", in *IEEE Global Communications Conference (GLOBECOM)*, December 2015.

4. H. Elshaer, C. Vlachos, V. Friderikos, and M. Dohler, "Interference-Aware Decoupled Cell Association in Device-to-Device based 5G Networks", in *IEEE Vehicular Technology Conference (VTC-Spring)*, May 2016.

5. C. Vlachos, H. Elshaer, J. Chen, V. Friderikos, and M. Dohler, "Bio-Inspired Resource Allocation for Relay-Aided Device-to-Device Communications", in *IEEE Vehicular Technology Conference (VTC-Fall)*, September 2016.

A.3 Posters

1. C. Vlachos and V. Friderikos, "Integrating Device-to-Device (D2D) Communications in 5G Networks", in *OFCOM's 5G and future technology event*, March 2015.

2. C. Vlachos, G. Chochlidakis, J. Heide, and V. Friderikos, "Network Coded Compression-based Caching for Device-to-Device Communications", in *European Conference on Networks and Communications (EuCNC)*, June 2016.